SVENSKA GEOFYSISKA FÖRENINGEN

VOLUME 9, NUMBER 3 **Tellus** AUG. 1957

A QUARTERLY JOURNAL OF GEOPHYSICS

Carl-Gustaf Rossby in memoriam

28 December 1898—19 August 1957

The death of Carl-Gustaf Rossby is a great loss for meteorology, oceanography and geophysics in general. One of the great pioneers and leaders in our science has left us. Carl-Gustaf Rossby was an unusual man. His exceedingly clear intellect was combined with a deep desire to serve science and mankind, a unique capability to stimulate his associates and a friendly and personal way of initiating and maintaining contact with them.

Rossby's contributions to science are mainly in the fields of dynamic meteorology and oceanography. Well-known concepts such as Rossby-waves, the Rossby-diagram, the Rossby-parameter and the Rossby-number will remain in the meteorological vocabulary for

years to come. His theoretical discussions were founded on a deep insight into the physics of the problems based upon a detailed knowledge of and feeling for the "real" atmosphere and ocean. He had unusual insight into the fundamental processes in nature even when obscured by unimportant details. His approach was basically heuristic. Brutal simplifications, to be justified afterwards, paved the way to the centre of the problems. A characteristic saying often used by him was: "The proof of the pudding lies in the eating." His imagination was constantly developing ideas which were submitted on trial to the group working around him, discussed, improved or rejected, but always stimulating the activity in the group, resulting in a strong feeling of the teamwork with the master. Such working sessions are unforgettable memories to all who worked in close association with him.

But Rossby's interest in a problem never stopped with having solved certain theoretical aspects of it. A strong feeling of responsibility to society urged him to carry it into operation. And much of his work also originates from his attempts to see the place of meteorology in society. Typical examples are his basic interest in weather forecasting, his attempts to couple meteorology and agriculture, in bringing forward the problems of atmospheric chemistry, his great interest in the deep sea circulation of the oceans both from the point of view of climatic changes and the discussion of radioactive waste and finally his concern about the hydrologic cycle in nature. Once a sound idea was born it was pursued with great consistency, first by himself, then gradually by letting more and more of it be taken over by his students.

Students were always attracted to Rossby resulting in an institute wherever he stayed for some time. The meteorology departments at Massachusetts Institute of Technology and University of Chicago as well as the International Meteorological Institute in Stockholm are his creations. His brave initiative also contributed greatly to the start of well-known scientific publications such as the Journal of Marine Research, the Journal of Meteorology and Tellus.

Rossby's interest in the planetary problems of meteorology made him very active in international collaboration. In particular after the last war and on moving to Sweden this internationalism absorbed his interest more and more. Many are the discussion groups he has brought together from countries all over the world and it is characteristic that he sought the contacts with individuals wherever he went rather than institutes and organizations they represented. In this way the meetings were informal and fruitful both scientifically and personally.

Professor Rossby will be especially remembered for his generosity, friendliness and sparkling humor. He took great interest in his students and often expressed the idea that you have to consider everyone of them as a special project to be developed according to the personality involved. His enthusiasm and optimism kept him spiritually young and it is characteristic that until the end it was a group of young people he gathered around him.

Rossby had the simplicity of the great. His life was his work. He died during a conference making great plans for the future. Carl-Gustaf Rossby is dead but his name and his spirit will be carried on by friends, students and colleagues all over the world.

B. Bolin.

The Threefold Structure of the Atmosphere and the Characteristics of the Tropopause

By FR. DEFANT and H. TABA, International Meteorological Institute in Stockholm

(Manuscript received March 25, 1957)

Abstract

All available soundings for January 1, 1956 have been studied with respect to their tropospheric and stratospheric temperature structure and with emphasis on the height and form of the tropopause. It was found that the latitudinal influence on the total vertical structure of the soundings was not dominating, but rather the location of soundings with respect to the two main westwind maxima (polarfront jet and subtropical jet) of the westerlies seemed to be important. A selection of soundings according to this principle (north of polarfront jet, between the two jets, south of subtropical jet) made it possible to distinguish between different types of soundings. This fundamental difference is discussed and will form the basis for future studies of world wide circulation changes. Furthermore the distribution of tropopause height and temperature has been mapped and studied on a hemispheric scale. Each of the above mentioned types possesses a typical vertical temperature structure and a characteristic tropopause heigth. Two rapid changes in tropopause height (tropopause breaklines) connected with the main westwind belts separate areas possessing an almost uniform tropopause level. Inside each of these areas a remarkable quasi-barotropic state of the upper tropospheric layers is shown.

Similar discussion has been devoted to the nature of soundings inside or close to the jets or the breaks in tropopause height. Here the soundings have a more complicated vertical structure and their interpretation becomes difficult.

Introduction

Until the present time synoptic and theoretical work in meteorology has been more or less restricted to the study of atmospheric conditions up to the 300 mb surface (about 9 km). This is especially true for the daily routine synoptic work. Although numerous studies have included the stratospheric conditions, these studies are mainly restricted to certain areas. Worldwide hemispheric investigations for stratospheric levels are extremely rare in meteorology. This is partly due to the fact that complete hemispheric data for levels above 300 mb are not available, and also because the contribution of stratospheric developments to that at lower levels has not generally been considered very important.

There are certain important features in the upper atmosphere (upper part of the troposphere and the stratosphere), which should be closely studied. We find at these levels the polarfront jet connected with the polarfront and the subtropical jet located mainly at about 30° lat. and at a height of about 200 mb. These two jets are the main windmaxima of the broad westerlies. While the polarfront jet has been studied rather thoroughly and its hemispheric existence has been shown, the subtropical jet has not been investigated enough. It is not shown yet, that the subtropical jet is a continuous belt around the hemisphere, however, climatological studies and certain vertical cross-sections in different meridional planes suggest its more or less continuous existence around the earth.

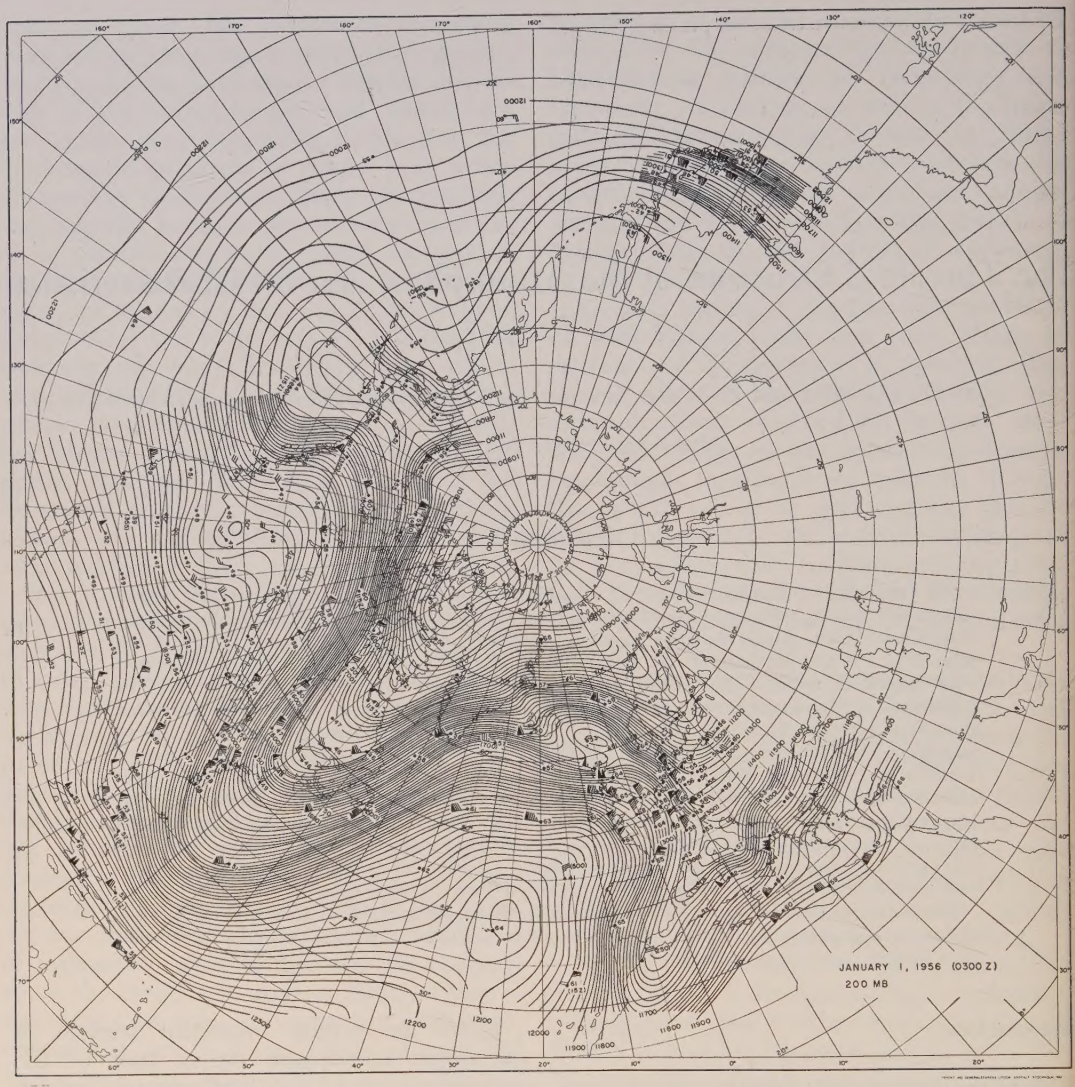


Fig. 1. 200 mb contour map of January 1, 1956 (0300 Z) (Thick contours for every 100 dyn. m., thin contours for every 20 dyn. m.; Temperatures in C° and minus sign omitted. Windspeed in kts.)

Another very important feature of the higher atmosphere is the variation of the tropopause height. The temperature field in the stratosphere has not either been studied enough.

The purpose of this paper is to investigate the behaviour of the atmosphere from the tropopause level upward. Special emphasis will be given to the variation of the tropopause height in vertical and in latitudinal direction. This is being done for a period beginning 1 January 1956. All available soundings of the northern hemisphere have been plotted and studied thoroughly. This period was selected as hemispheric maps from the surface up to the 300 mb level showed certain periodic and pronounced changes of the world-wide patterns. This paper deals only with January 1, 1956. During the preparation of the material, it was found that the Russian and Chinese radiosondes did not extend above 300 mb, so that no, or only little information was available in these areas. The analysis covers therefore the Pacific Ocean, North America, the Atlantic Ocean and Europe.

1. General synoptic situation on January 1, 1956

It is clear that the 200 mb level is most suitable for identifying the subtropical jet, as the highest wind speed of this phenomenon occurs close to this level. The polarfront jet can also be noticed in this level as the wind velocity remains still strong and the gradient of the contour lines shows a well marked concentration at the 200 mb level (upper polarfront). A further advantage is the well known reverse of the meridional temperature gradient, so that the temperature field is also a guide in distinguishing the two separate wind belts. This will be shown in detail later. As an exampel Fig. 1 gives the situation on January 1, 1956 (0300 z). This analysis clearly shows a strong meandering circulation pattern. Particularly strong troughs can be noticed just off the American East Coast, over Europe and over the Japanese Islands. A weaker trough appears south of Alaska. The wind values over the North American Continent together with the gradient of the contour lines show the upper polarfront jet running across Northern Canada and going around the trough at the East Coast of America. At the same time, the high wind velocities over the Southern United States and the Caribbean Islands show the subtropical jet quite well. Over the Atlantic the polarfront jet seems to run from New Foundland to the southern part of Greenland, to the north of Iceland and into the middle of Europe.

The subtropical jet in this area continues from the Caribbean Islands towards the northeast, crosses the northern part of the Atlantic Ocean and enters England. Over France the two jets nearly join, running southwest to the African West Coast and show up again over North Africa.

Fig. 1 also shows the reversed temperature gradient at 200 mb. North of the upper polarfront relatively high temperatures appear on the chart (sometimes higher than – 50° C). This indicates that in the area surrounded by the upper polarfront the tropopause is below the 200 mb level and the temperatures belong to the polar stratosphere. South of the upper polarfront temperatures between – 50 and – 60° C appear, where the tropopause is still below the 200 mb surface. Going further south the temperatures become relatively low, indicating that the tropopause intersects the 200 mb level and slopes upward to higher altitudes. Still further south the temperatures rise slightly and refer now to the higher troposphere. This can best be seen over the Atlantic area where the subtropical jet runs exceptionally far north. In this area temperatures lower than – 60° C are observed, indicating that most of the Atlantic is covered by a rather cold air of subtropical origin with the tropopause above 200 mb. Here the temperatures belong to the higher part of the troposphere.

This discussion indicates a threefold structure of the tropopause height which will be discussed in the following.

2. The tropopause map of January 1, 1956

To study this threefold structure in tropopause height, it was decided to prepare separate hemispheric tropopause maps. For this purpose about 200 soundings were plotted and checked for every day and for a period of seven days. The height of the tropopause was computed for each individual sounding. For most of the soundings it was easy to select the tropopause level by applying the standard method. (Either a noticable change of tropospheric lapse rate to an isothermal layer or to an increase of temperature with height.) Sometimes it was difficult, however, to find a well defined tropopause. This was almost always true for soundings, which met either the polarfront jet or the subtropical jet close to their wind maxima. In case of soundings running through the polarfront jet no distinct tropopause could be found. In the case of soundings going through the subtropical jet, two different breaks in lapse rate appeared, suggesting a double structure of the tropopause. Both of these cases will be discussed later.

Fig. 2 shows the tropopause map of January 1, 1956 (0300 z). In this map the two separate west wind belts stand out clearly in the gradient of the contour lines of the tropopause level. Two different bands picturing the break in the tropopause height connected with the existence of the polarfront—and the subtropical wind maxima can be seen all around the hemisphere. This double break in tropopause separates, in a clear manner, three differ-

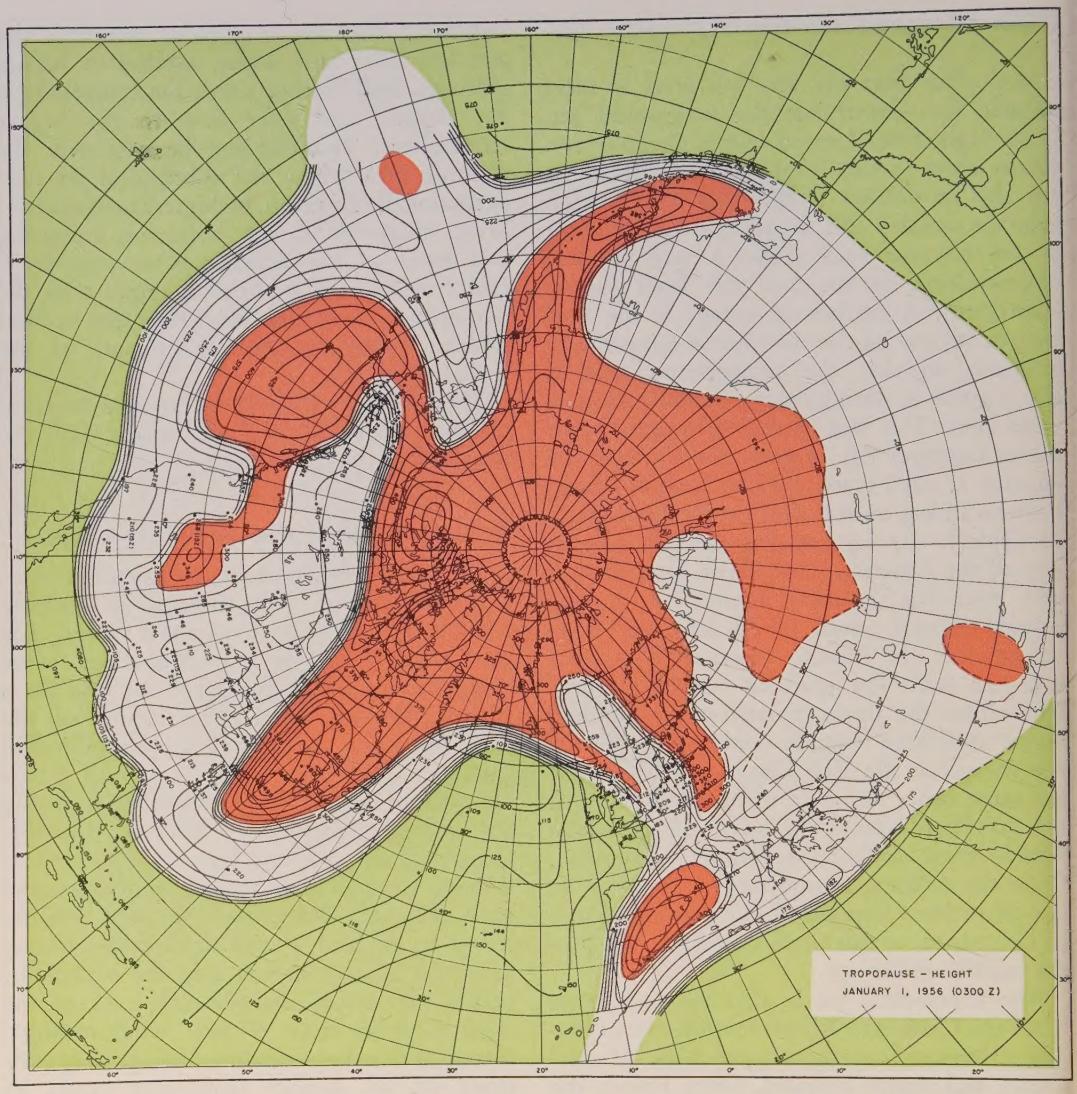


Fig. 2. Hemispheric map of tropopause height (in mb) of January 1, 1956 (0300 Z). Breaklines of tropopause height are shown by concentrations of isolines. Dashed lines over Asia give approximate continuation of the breaklines of tropopause height.

ent areas with different characteristic tropopause heights.

South of the subtropical break line the height of the tropopause is mostly around 100 mb or even higher. Only in those parts where this southern breakline goes far towards the north (in this map over the Atlantic Ocean) does the height of the tropopause drop down to about 175 mb, which can still be considered an exceptionally high tropopause at these latitudes according to synoptic experience.

In the area between the two breaklines the tropopause height stays around 250 mb. It is rather astonishing that the height of the tropopause does not vary by more than 50 mb inspite of the fact that the stations are located in widely differing latitudes. This can best be seen over the Northern American Continent where the two break lines are far apart.

In the area north of the polar break line, the tropopause heights range between 300 and 400 mb and in certain locations even as

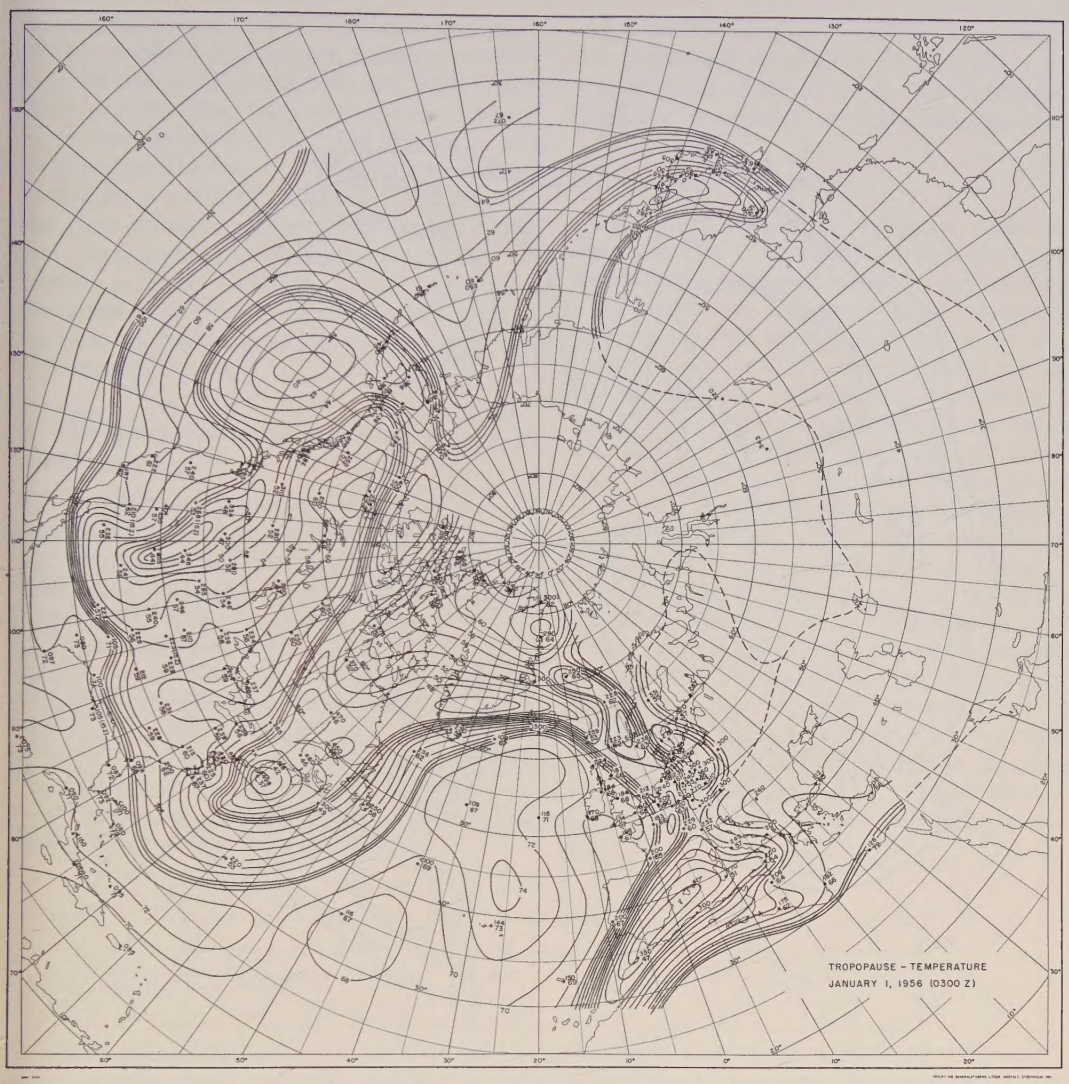


Fig. 3. Hemispheric map of tropopause temperature of January 1, 1956 (0300 Z). (For each station tropopause height in mb (three figures) and tropopause temperatures in C° (two figures) are plotted. Minus sign for temperatures omitted.) Note the striking similarity with Fig. 2.

low as 500 mb. This low tropopause height not only shows above cold polar vortices, but can also be found above cold polar outbreaks in relatively low latitudes; see e.g. the troughs off the East Coast of America, to the south of Alaska, over Northern Japan or even the weak trough east of the Rocky Mountains.

The tropopause map has the advantage to show in a clear manner the distribution of cold and warm tropospheric air masses, together with the location of both jets. One can say that the tropopause map gives an immediate three-dimensional picture of the state of the atmosphere. In this connection one should pay attention to a remarkable feature of this map, namely the location of the subtropical breakline over the Atlantic Ocean. Here the subtropical jet has been displaced far towards the north. The discussion of this feature will be left for a separate paper.

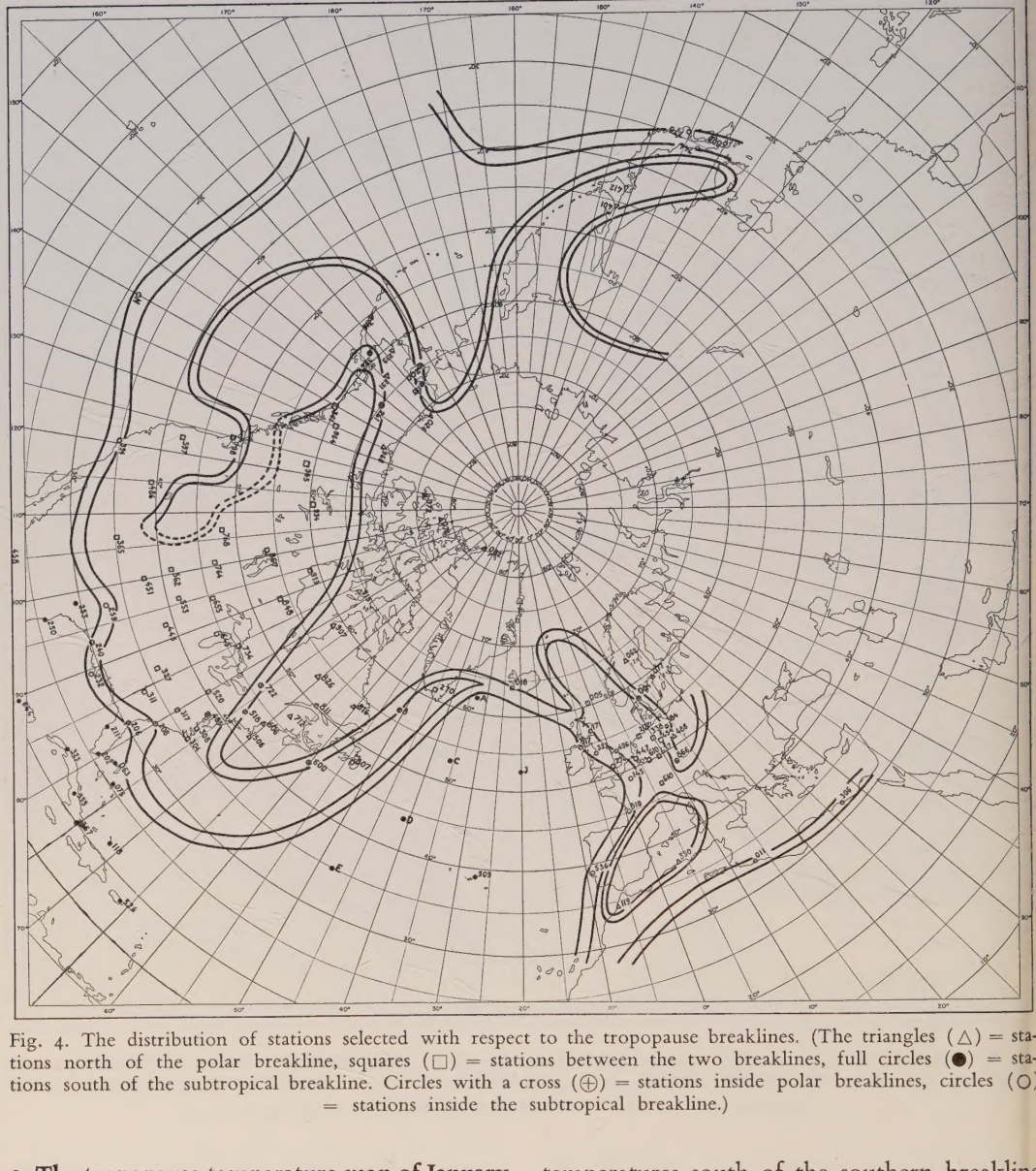


Fig. 4. The distribution of stations selected with respect to the tropopause breaklines. (The triangles (△) = stations north of the polar breakline, squares (□) = stations between the two breaklines, full circles (●) = stations south of the subtropical breakline. Circles with a cross (⊕) = stations inside polar breaklines, circles (○) = stations inside the subtropical breakline.)

3. The tropopause temperature map of January 1, 1956

Fig. 3 shows a similar analysis of the tropopause temperature. In comparing the pattern of this temperature field with the pattern of the tropopause height presented in Fig. 2, we notice a marked similarity. The breaklines in tropopause height and strong concentrations in the distribution of the tropopause temperature agree closly in location. The tropopause temperatures south of the southern breakline remain almost constant between – 67 and – 74° C. The same is true for the middle area between the two break lines where the tropopause temperature ranges between – 56 and – 60° C. North of the northern breakline the tropopause temperatures are mostly higher than – 50° C and in deep depressions of tropopause height (e.g. at the American East Coast) may be as high as – 37° C. This kind of temperature analysis offers a rather clear

distinction between three very different parts of the northern hemisphere which are separated by the main wind belts or, what is the same, by the breaklines in tropopause height or by strong bands with rapid change in the tropopause temperature. With the exception of the very northern latitudes (75 to 90° N) the close relation between tropopause height and tropopause temperature seems to be almost perfectly fulfilled in the sense that an increase of height goes parallel with a decrease in temperature or vice versa (compensation principle).

4. The presentation of different groups of soundings

The threefold structure in the tropopause heigth mentioned above will now be shown by presenting different series of soundings selected according to their location with respect to both breaklines. Fig. 4 shows the distribution of the stations selected. The stations are divided into five groups:

a) Stations north of the polar breakline (polar group of soundings)

b) Stations between the two breaklines (middle group of soundings)

c) Stations south of the subtropical breakline (subtropical group of soundings)

d) Stations close to or in the polar breakline (soundings through polarfront jet)

e) Stations close to or in the subtropical breakline (soundings through subtropical jet)

a) *Polar group of soundings*

The polar soundings and their mean are shown in *Fig. 5*. In these soundings the tropopause usually is formed below the 300 mb level. The tropopause height is more changeable than for the two types b and c discussed later. In some cases (near the center of polar vortices and inside deep troughs) the tropopause is lower than 400 mb, sometimes even below 500 mb (see e.g. No. 3, 15 and 16, Mould Bay, Portland and Nantucket). In the lowest layers of the troposphere a nearly isothermal structure or strong surface inversion is found. The higher tropospheric layers possess a nearly uniform lapse rate and in the stratosphere an almost constant temperature with height or even a slight decrease (above 200 mb) is observed. The total structure (troposphere plus stratosphere) is similar in almost all the soundings. This typical form of the polar soundings is best seen in the mean sounding on the right side of Fig. 5. *Table I* presents vertical lapse rates of the mean polar sounding.

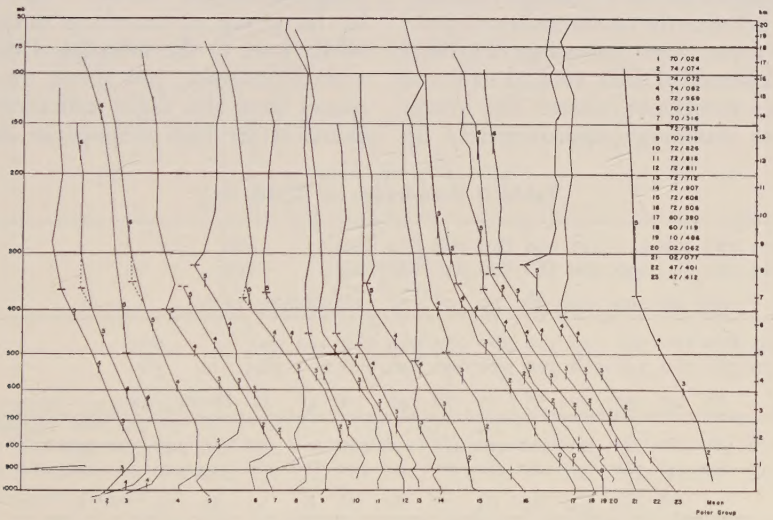


Fig. 5. The vertical temperature structure of soundings north of the polar breakline (polar group). Note the location of the tropopause below 300 mb. The — 10°, — 20°, — 30° C a.s.o. are indicated by 1, 2, 3 resp. Mean temperature distribution of these soundings is given on the right side.

Table I (hundredth of °C/100 m)

layer (mb)	from	940	920	900	880	860	840	820	800	780	760	740	720			940
	to	920	900	880	860	840	820	800	780	760	740	720	700			700
		22	26	14	22	18	18	24	22	24	35	38	39	Mean		25
layer (mb)	from	700	680	660	640	620	600	580	560	540	520	500	480			
	to	680	660	640	620	600	580	560	540	520	500	480	460			
		41	61	65	55	67	59	81	71	78	70	58	58			
layer (mb)	from	460	440	420	400	380										700
	to	440	420	400	380	360										360
		40	55	54	35	39								Mean		61
layer (mb)	from	360	340	320	300	280	260	240	220	200	180	160	140	120		360
	to	340	320	300	280	260	240	220	200	180	160	140	120	100		100
		00	05	10	05	08	04	08	02	01	08	10	01	00	Mean	05

From Table I one may see that the lower polar troposphere up to 700 mb is rather stable (mean lapse rate – 0.25° C/100 m) while the upper troposphere tends towards a more unstable stratification (mean lapse rate – 0.6° C/100 m, sometimes – 0.7 to – 0.8). The total mean tropospheric lapse rate is – 0.46°/100 m. In the polar stratosphere almost isothermal conditions are present (mean lapse rate – 0.046°/100 m). The mean tropopause height is 370 mb (approx. 8 km) and the tropopause temperature is – 49° C.

b) *Middle group of soundings*

Fig. 6 a, b shows all the soundings from stations situated between the two breaklines together with their mean. These soundings in general look rather different in their vertical structure from the ones mentioned above. The lowest parts of these soundings, approximately up to 850 mb, vary considerably from one sounding to another. Often strong surface inversions are observed which partly are low level polarfront intersections. The higher tropospheric layers up to the tropopause show a surprisingly uniform vertical temperature distribution. The tropopause is found at about 240 mb and in no case seems to be above 200 or below 300 mb. At the tropopause the nearly constant vertical lapse rate of the higher troposphere changes to a more or less constant temperature with height in the stratosphere. However in the higher stratospheric layers (from 150 mb upwards) the temperatures are mostly – 60° C and lower. In this group of soundings only those somewhat closer to the subtropical breakline show a small inversion just above the main tropopause. Here also slight indications of an additional rather high tropopause above 100 mb

Table II (hundredth of °C/100 m)

layer (mb)	from	980	960	940	920	900	880	860	840	840												980		
	to	960	940	920	900	880	860	840	820	800												860		
		00	00	00	00	00	00	00	12	16	Mean	14												
layer (mb)	from	800	780	760	740	720	700	680	660	640	620	600										800		
	to	780	760	740	720	700	680	660	640	620	600	580										580		
		19	40	40	40	55	75	76	40	70	48	52	Mean	52										
layer (mb)	from	580	560	540	520	500	480	460	440	420	400	380	360	340	320	300	280	260						580
	to	560	540	520	500	480	460	440	420	400	380	360	340	320	300	280	260	240						240
		72	78	70	73	69	63	73	80	86	83	73	85	75	76	61	62	40	Mean	72				
layer (mb)	from	240	220	200	180	160	140	120	100	80	60											240		
	to	220	200	180	160	140	120	100	80	60	50											50		
		+18	+11	00	15	24	19	16	01	+06	+03	Mean	12											

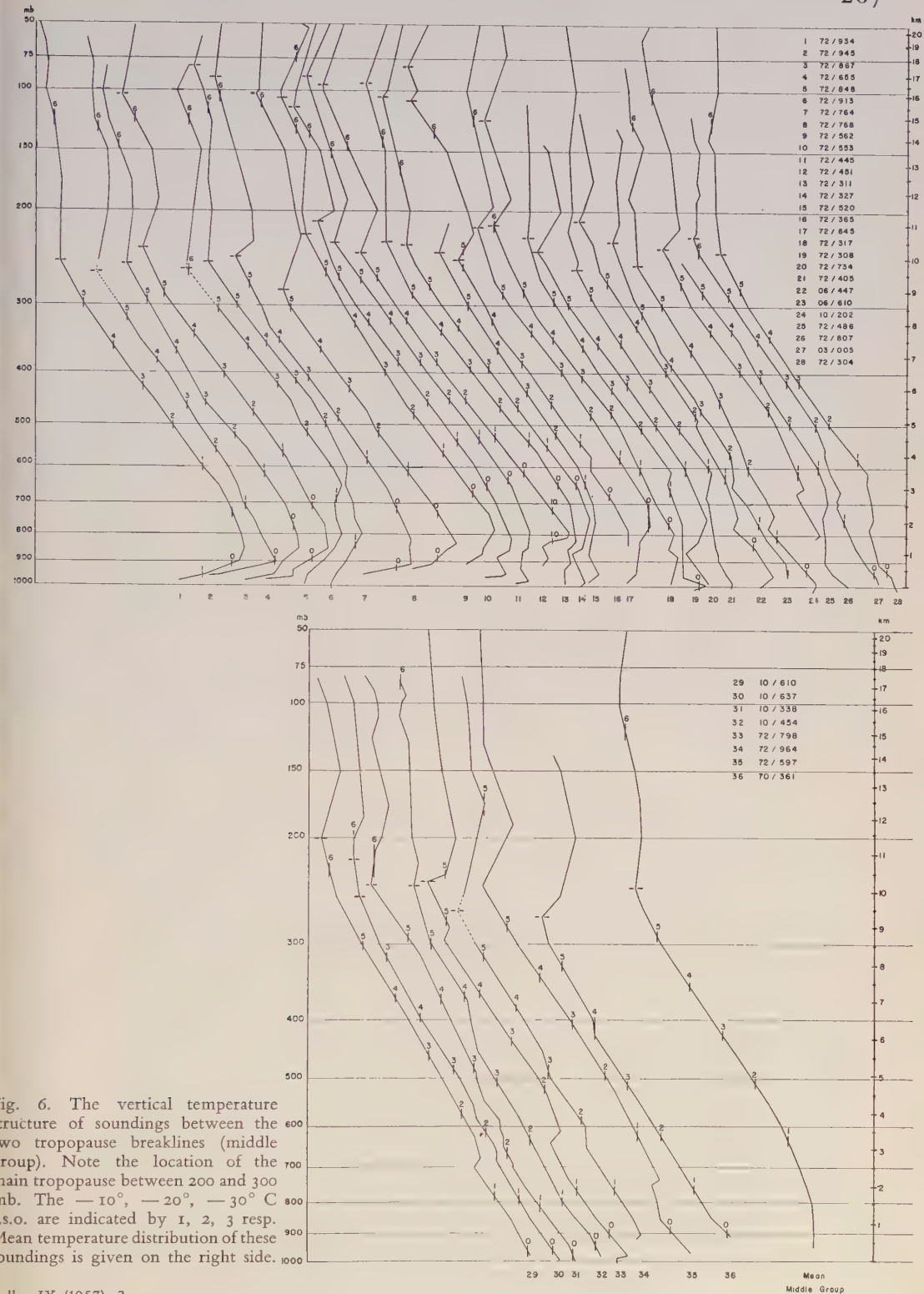


Fig. 6. The vertical temperature structure of soundings between the two tropopause breaklines (middle group). Note the location of the main tropopause between 200 and 300 mb. The — 10°, — 20°, — 30° C a.s.o. are indicated by 1, 2, 3 resp. Mean temperature distribution of these soundings is given on the right side.

can be noticed (e.g. No. 7, 9, 10, 11, 12, 16; Bismark, North Platte, Omaha, Columbia, Dodge City, Albuquerque).

Table II presents vertical lapse rates for different height intervals of the mean sounding of this group.

The lowest layers show rather variable conditions and the mean lapse rate of − 0.14° C/100 m is only a result of the averaging process. Above 800 mb the lapse rate of the mean middle sounding increases with height up to about 340 mb and decreases slightly until the tropopause is reached (240 mb). A total mean lapse rate of the troposphere is − 0.64°/100 m. The tropopause temperature is − 56° C. Even in the mean sounding a slight inversion appears above the tropopause and from 170 mb upwards the temperature decreases only very little with height (− 0.12° C/100 m) up to 90 mb, where a secondary tropopause level can be found. From there on the temperature remains nearly constant.

c) *Subtropical group of soundings*

Fig. 7 shows all the soundings from stations south of the subtropical jet, together with their mean sounding. These soundings possess again a rather similar vertical temperature structure. Even in the lower layers the differences in temperature are rather small. The 0° C level is around 620 mb (approx. 4.0 km). The tropopause appears in all soundings at rather high altitudes (near the 100 mb level). The vertical lapse rate seems to be quite uniform in the whole troposphere and at the tropopause level temperatures of − 70° C or lower are observed. A sharp break in lapse rate appears at the tropopause and from there on the temperature increases with height up to 50 mb. At 50 mb the observed temperatures are mostly around − 60° C. There seems to be no indication of a constant temperature with height in the stratosphere, at least not in the layer between 100 and 50 mb. Above 50 mb the observational evidence is rather poor, therefore nothing conclusive can be said about it. A study of the mean sounding shows the above described structure in a clear manner. Table III presents vertical lapse rates for different layers of this mean sounding.

In the lower troposphere (1,000—600 mb) where the temperature varies more from sounding to sounding, the mean vertical lapse rate is − 0.46° C/100 m and from there on a rather constant lapse rate of − 0.73° C/100 m extends upwards to 250 mb. At this level the lapse rate changes a little and the mean value from 250 mb to the tropopause is about the same as in the lowest layer (− 0.41° C/100 m). In the subtropical stratosphere the mean

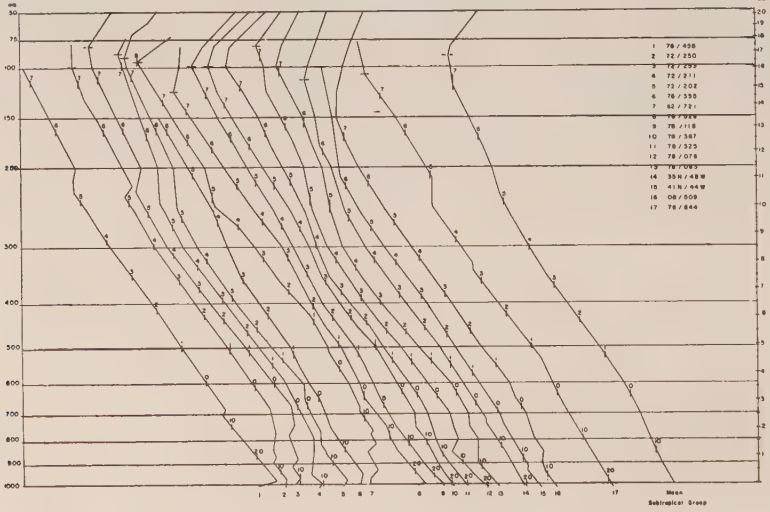


Fig. 7. The vertical temperature structure of soundings south of the subtropical breakline (subtropical group). Note the high level tropopause (above 150 mb). Location of — 10°, — 20°, — 30° C a.s.o. are indicated by 1, 2, 3 resp. Mean temperature distribution of these soundings is given on the right side.

Table III (hundreth of °C/100 m)

layer (mb) from	980	960	940	920	900	880	860	840	820	800	780	760	740	720					
to	960	940	920	000	880	860	840	820	800	780	760	740	720	700					
	32	48	40	50	50	44	50	30	40	30	41	50	32	55					
layer (mb) from	700	680	660	640	620	600													
to	680	660	640	620	600	580													
	43	48	56	52	56	72													
layer (mb) from	580	560	540	520	500	480	460	440	420	400	380	360	340	320	300	280	260		
to	560	540	520	500	480	460	440	420	400	380	360	340	320	300	280	260	240		
	57	72	73	67	87	66	100	71	81	76	69	72	73	58	82	58	73		
layer (mb) from	240	220	200	180	160	140	120												
to	220	200	180	160	140	120	100												
	47	39	41	60	46	44	10												
layer (mb) from	100	90	80	70	60														
to	80	80	70	60	50														
	17	+ 13	+ 50	+ 33	+ 30														

980
580
Mean 46

580
240
Mean 73

240
100
Mean 41

100
50
Mean 27

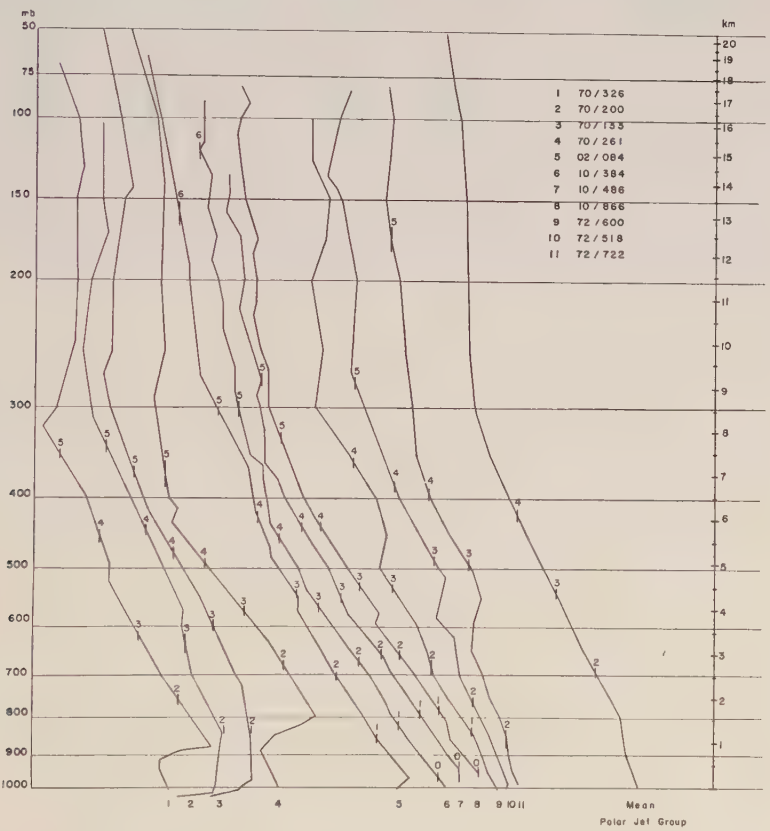


Fig. 8. The vertical temperature structure of soundings inside the polar breakline. The tropopause is not well defined and the temperature continues to decrease upward. The mean of these soundings is presented on the right side of the figure. The — 10°, — 20°, — 30° C a.s.o. are indicated by 1, 2, 3.

vertical increase of temperature with height was computed to be $+0.27^\circ$ C/100 m. The mean tropopause is a little above 100 mb surface (exactly at 90 mb) and the tropopause temperature is -72° C.

In the area south of the subtropical breakline a typical sounding shows therefore a more or less uniform temperature decrease in the whole troposphere, which extends upwards to higher altitudes (near to 100 mb). From there on a noticeable increase of temperature with height is observed. The tropopause level is in most cases well defined. It must be clearly understood that soundings farther south would show the tradewind inversion. In addition the tropopause would show up in still higher altitudes (between 90 and 60 mb) and the tropopause temperature would still decrease (in the equatorial region even as low as -80° C, see e.g. No. 7, Khartoum).

d) *Polar jet group of soundings*

In soundings which intersect the wind maximum connected with the polarfront (Fig. 8) no clear tropopause is observed. The temperature decreases mostly with height both in tropo- as well as in stratosphere. This can be seen in the soundings of Berlin, Munich, Dresden and Albany (No. 6, 8, 7 and 10). In some of these soundings slight tropopause indications may be seen. The form of these soundings and the breaks in lapse rat, depend very much on the exact location of the sounding with respect to the polar jet and on the steepness of the upper polarfront. One has to be rather careful not to mix the tropopause with breaks in lapse rate appearing when the sounding passes the polarfront layer. The mean sounding also shows the continuous temperature decrease with height. A broad layer with a slight change in lapse rate separates the troposphere from the stratosphere.

e) *Subtropical jet group of soundings*

In Fig. 9 soundings through the subtropical jet are presented. In these soundings it is also difficult to select a single tropopause. But contrary to the case of polar jet soundings, two different tropopause levels are generally present. From these, the lower one is formed around 200 mb level or a little above, while the higher one shows up at about 100 mb. This twofold tropopause structure can clearly be seen in the mean sounding. Immediately after passing the lower tropopause, a frontlike temperature inversion can be noticed in most cases. This lower tropopause seems to be closely related to the level of highest wind speed of the subtropical jet. Above this inversion the temperature falls off again sharply

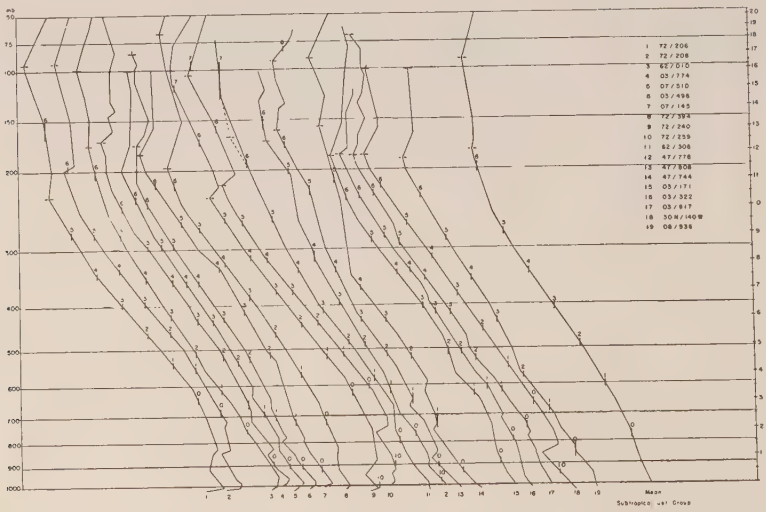


Fig. 9. The vertical temperature structure of soundings inside the subtropical breakline. Note the double tropopause level. The -10, -20, -30° C a.s.o. are indicated by 1, 2, 3. The mean of these soundings is given on the right side of the figure.

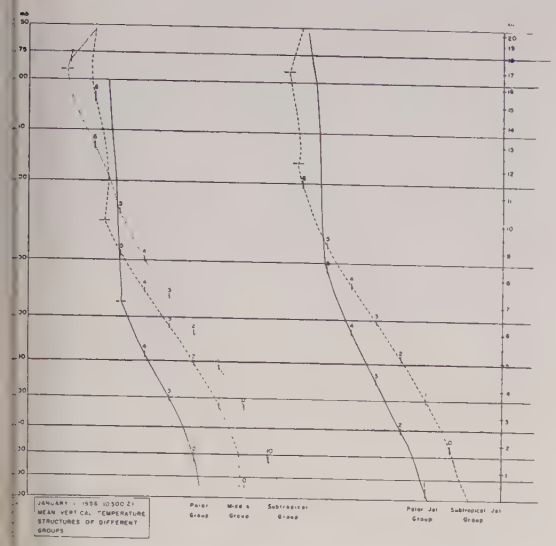


Fig. 10. Mean vertical temperature distribution fo the five different groups. The — 10, — 20, — 30° C a.s.o. are indicated by 1, 2, 3.

and at the higher tropopause this decrease changes abruptly to an increase with height. Here also it depends critically on the exact location of the sounding with respect to the subtropical jet maximum whether the lower or the higher tropopause is dominating.

5. Comparison of the mean soundings

Fig. 10 presents the five mean soundings side by side. From this picture it is rather obvious that besides the strong baroclinicity in the troposphere between polar and middle soundings concentrated mainly in the polarfront zone an equally important baroclinicity is present in the stratosphere between middle and subtropical group, except that the horizontal temperature gradient is reversed. However, the stratospheric differences in temperature between polar and middle group are rather small. For a circulation type such as that of January 1, 1956, where the subtropical jet

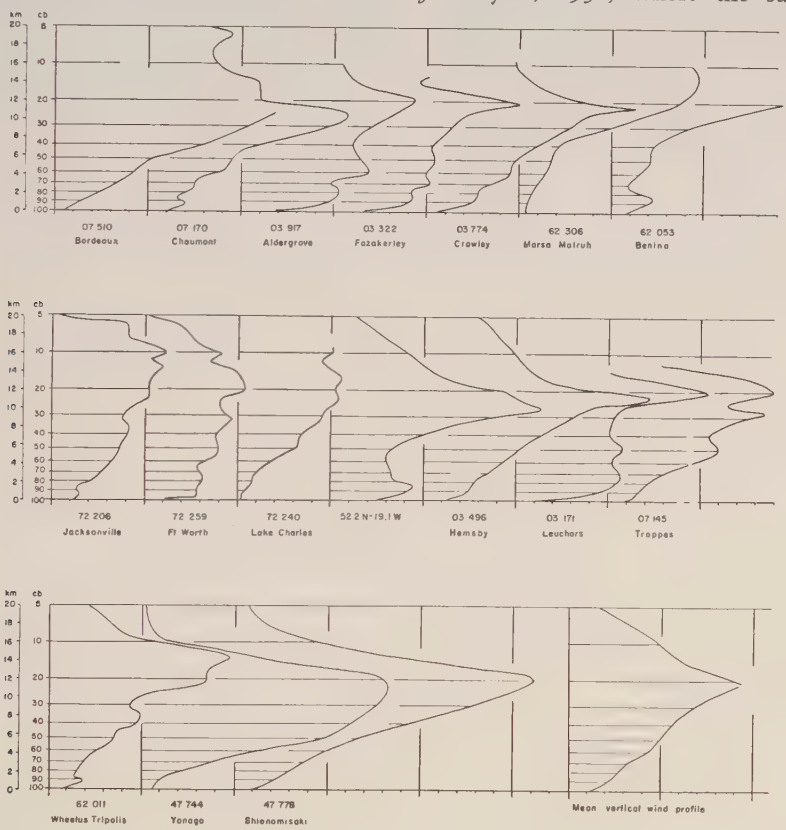


Fig. 11. Vertical wind profiles of January 1, 1956, from stations close to the subtropical tropopause breakline. The mean vertical wind profile is presented on the right side of the lowest row. Each small interval on the abscissa indicates 10 kts and the main vertical lines intervals of 50 kts.

remains mainly near latitude circle 30° N and the polar jet is mostly far apart the two baroclinic zones are also well separated. When the subtropical system moves northward, as e.g. over the Atlantic Ocean, then the tropospheric and the stratospheric baroclinicity are nearly superimposed. In this case the middle part disappears and the whole baroclinicity exists between subtropical and polar group in tropo- and stratosphere, but in the stratosphere in the reversed sense.

6. Vertical wind profiles of subtropical jet region

In the foregoing discussion it has been mentioned that the obvious breaks in tropopause height are closely related to the main wind belts. To show this more clearly the vertical wind profiles of stations close to, or in, the southern breakline of the tropopause (group e) are presented in Fig. 11. From these vertical profiles it is rather evident that the subtropical breakline in tropopause height is closely related to the subtropical jet, which in all of these profiles appears around the 200 mb level. In some of these examples a broad wind maximum can be noticed (Jacksonville, Ft. Worth and Lake Charles), that is when the sounding did not hit the jet center. There are also some examples of extremely sharp wind maxima in the vertical, indicating that the sounding met the center (e.g. Crawley, Liverpool (Fazakerley), Hemsby, Shionomisaki a.o.). The vertical wind profiles of Liverpool and Crawley are rather beautiful examples of a well defined jet restricted to a layer of a small vertical extent of about 50 mb (1.5 km). There exists a strong vertical wind shear above and below the jet center.

By making a mean profile from these 20 stations, the phenomenon of the subtropical jet in its hemispheric extent is well shown. That this jet phenomenon is not connected with the polarfront can be seen from the temperature distributions presented in Fig. 9.

It was not possible to present a similar wind profile for the polar breakline of tropopause height, as stations over Canada did not have good enough wind information and over the Atlantic the two jet belts were so close together that both phenomena showed up in vertical profiles.

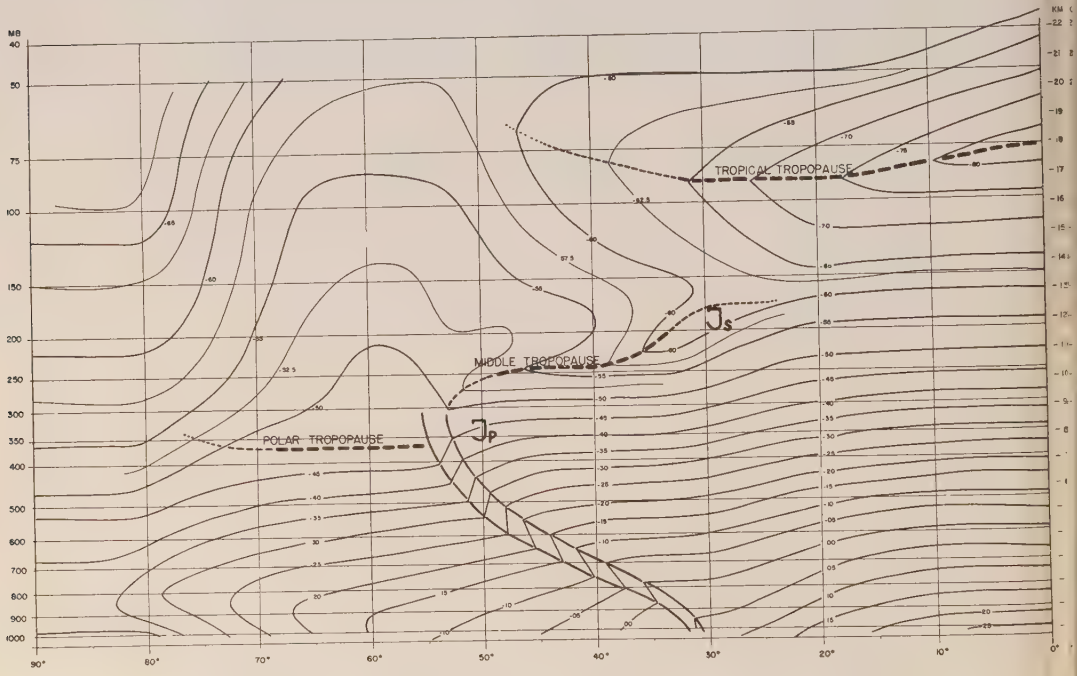


Fig. 12. Mean meridional cross-section of actual temperature for January 1, 1956. Note the threefold structure of the tropopause. I_p = location of polarfront jet, I_s = location of subtropical jet.

7. Summary of the foregoing discussion and presentation of a mean meridional cross-section

In Fig. 12 a mean meridional cross-section based on the mean values of temperature obtained above is presented. The purpose of this figure is to give the reader a general picture of the important features of the temperature field of the atmosphere in meridional direction. The threefold structure of the tropopause can easily be seen.

The foregoing discussion summarized in Fig. 12 divides the atmosphere from the polar to the tropical region into three widely differing parts represented by their mean soundings. Each part has its own typical vertical temperature structure in troposphere and stratosphere together with a characteristic tropopause height.

The polar and the middle part are separated in the troposphere by the polarfront, which extends upwards to about 300 mb. There exist a definite break in the tropopause height related to the polarfront jet (polar breakline in tropopause height; mostly between 250 and 350 mb). As mentioned before the strongest baroclinicity of the troposphere is concentrated in the inclined polar frontal zone. The tropospheric meridional temperature structure of the middle part can be considered almost barotropic. This is true especially for the upper troposphere.

The middle and the subtropical, tropical part resp., are not separated by any frontal surface in the troposphere. However, there exists a small baroclinicity between the almost quasibarotropic stratification of the middle and subtropical troposphere. This is indicated by the inclination of the isotherms in Fig. 12 around lat. 25—30° N. On top of this weak baroclinic zone another break in the tropopause is present (subtropical breakline in tropopause height; mostly between 240 and 100 mb). The vertical extent of this breakline is much larger than that of the polar breakline. The middle tropopause seems to continue southward and increases its height to about 180 mb. The subtropical jet is located in the mean just beneath this continued part of the middle tropopause. This is why the soundings through the subtropical jet indicated in all cases a lower tropopause as well as the higher subtropical one. The subtropical tropopause has an average height of about 90 mb and seems to rise towards the equator. A close study of the middle soundings indicates that the subtropical tropopause continues also towards north, increases in height and gradually disappears.

So far we have not mentioned anything about the temperature field in the stratosphere. A study of the Fig. 12 shows that the stratospheric temperature difference between the polar and the middle part is not very pronounced. However, this difference becomes quite noticeable between the stratosphere of the middle part and the upper subtropical troposphere. In ordinary cases when the two main westwind maxima are separated by the middle part the stratospheric meridional temperature gradient is not very pronounced. But whenever the subtropical system is displaced northward and the middle part disappears the total meridional temperature difference of the stratosphere concentrates in a frontlike zone above the polarfront but with an opposite inclination.

From this discussion we may now easily understand why a hemispheric presentation of tropopause height as well as tropopause temperature makes it possible to visualise the characteristic features of the three dimensional structure of the atmosphere. The study of the atmospheric field of temperature by means of certain isobaric levels can lead to serious misinterpretations. It is also rather difficult to draw any conclusion in this way.

In this discussion we have so far excluded the area from about 80 to 90° N (inner polar cap). This region is affected by different factors such as advection in different layers and from different directions or radiational influences. Here no representative mean state can be found. This does not mean that we underestimate the importance of this part of the atmosphere.

Fig. 13 presents the meridional distribution of potential temperature of the section shown in Fig. 12. In the troposphere north of the polarfront the isentropic surfaces rise towards the pole with an almost constant inclination. The concentration of these surfaces starting at the surface around lat. 65° and continuing towards the pole up to a height of about 650 mb is an indication of the so called arctic front. The strongest concentration can be seen inside

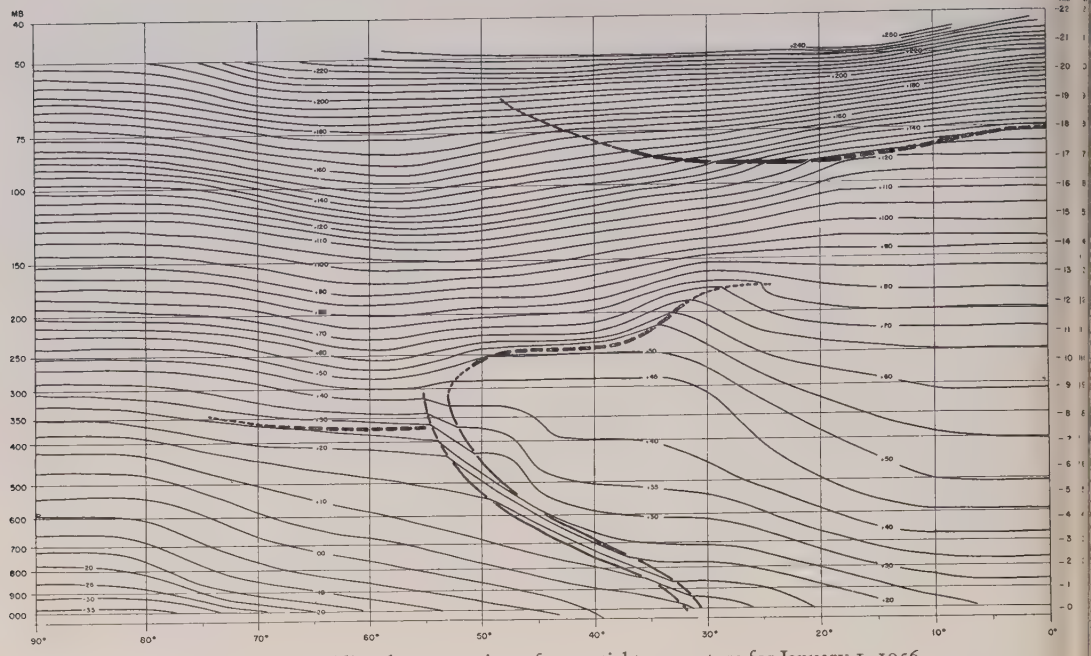


Fig. 13. Mean meridional cross-section of potential temperature for January 1, 1956.

the boundaries of the polarfront. The sharp slope of the isentropes immediately south of the polar jet (about lat. 45°) seems to be closely related to the existence of the jet. Between 45 and 30° N the isentropic levels are almost horizontal. This is an indication of the quasibarotropic state of the middle part of the atmosphere described before. From 30° southward to about 20° N a further downward slope of the isentropes indicates the difference between middle and subtropical air. Further south the atmosphere is again quasi barotropic. Here the concentration of the isentropes especially in the upper troposphere has a close relation to the existence of subtropical jet near 200 mb.

In the upper troposphere of the tropical region between the equator and 20° N the stratification is again barotropic. Above the tropical tropopause (tropical stratosphere) and in the middle stratosphere (from about 20° N) the isentropes slope downwards until about 65° N. From there on towards the pole they incline upwards. In the space where the middle tropopause rises and gradually disappears (northern side of subtropical jet) a frontlike concentration of the isentropes can be seen. Whenever the subtropical jet moves far northward and comes close to the polar jet,

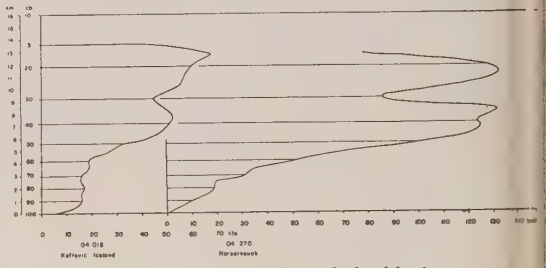


Fig. 14. Examples of complicated double jet structures in the vertical wind profiles of Keflavik (Iceland) and Narsarssuak (Greenland).

this stratospheric frontal slope becomes more pronounced and may be interpreted as a stratospheric continuation of the polarfront. In such cases the two jet maxima approach and nearly superimpose, resulting in a double wind maximum in the vertical wind fields. Fig. 14 shows an example of such a double jet over southern Greenland and Iceland where the two jets are close to each other (see also Figs. 2 and 3).

Acknowledgements

The authors wish to express their thanks to Mrs. Birgitta Lorensson and Mrs. Birgit Dahlborg for helping with the preparation of the material and the figures.

Uncertainty of Initial State as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns

By PHILIP DUNCAN THOMPSON, Lt. Colonel, U. S. Air Force, Joint Numerical Weather Prediction Unit, Washington, D.C.

(Manuscript received February 5, 1957)

Abstract

This article deals with the predictability of the atmosphere or, more exactly, with the gradual growth of "inherent" errors of prediction, due to errors in an initial state that is reconstructed from measurements at a finite number of points. By investigating the initial time-derivatives of the error arising from *random* analysis error, it is found that the increase of the RMS (root-mean-square) wind error in predictions over periods of a few days depends on:

1) the period of the forecast
2) the *initial* RMS vector wind error
3) the difference between the characteristic scale of the initial error field and the scale of fluctuations in the true initial flow pattern
4) the area average of the vertical wind shear between 250 and 750 mb
5) the RMS vector deviation of the wind at about 500 mb from its area average and
6) the average static stability of the atmosphere.

The joint effect of these various factors is given explicitly by a single equation, relating the increase of inherent error to the statistical properties of the initial error field and true initial flow pattern.

In many winter situations, and for initial error fields whose scale is typical of the present observational network, the inherent RMS vector wind error may double its initial value after two days, and rise to the error of sheer guessing in about a week. Doubling the overall density of regular reporting stations would virtually eliminate the increase of *inherent* error in forecasts over a few days. It is also found that zonally-averaged wind fields are inherently more predictable than unaveraged wind fields, at least in cases of predominantly barotropic flow.

The results outlined above are interpreted in the terms of a variety of practical and administrative problems, in each of which a dominant factor is the predictability of the atmosphere. Examples are the problems of estimating limits of confidence in forecasts, deciding on the most economical density and distribution of regular reporting stations, fixing the maximum range beyond which detailed "forecasts" have lost essentially all predictive and economic value and, finally, that of establishing the point of rapidly diminishing returns in the development of more complicated and costly methods of prediction.

I. Introduction

One of the most far-reaching questions in meteorology concerns the "predictability" of the atmosphere—i.e., not merely the extent to which its behavior is predicted in practice, but the extent to which it is possible to predict it with a theoretically complete knowledge of the physical laws that govern it. On the purely scientific side, it is an important point of doctrine to know whether or not our uncertainty as to the atmosphere's future state is accounted for by economic (or, ultimately, human) incapacity to observe and compute, or whether it is essential and due to some irreducible minimum

of indeterminacy that lies beyond human limitation.

On the practical side, common experience indicates that incomplete data coverage leads to a rapid and widespread decay of predictability over periods of more than a day or two, and it is generally granted that increasing the density of regular reporting stations would result in a qualitative increase in predictability and in the economic value of prediction. The cost of (say) doubling the density of upper air stations, however, would be enormous—certainly on the order of hundreds of million U.S. dollars. Thus, the question of increasing data density is an economic problem of national or international proportions, and it is hardly surprising that national weather services in the past have been reluctant to double their budgets for a problematical return in the accuracy of forecasts. Until one can quantitatively predict the predictability that would result from increased data coverage, it is impossible to establish the point of rapidly diminishing returns, beyond which further outlay would be unprofitable. That, of course, is the real problem.

Another related and equally mundane question is whether or not it is useless to try to predict the state of the atmosphere in detail over periods longer than several days. It is often asserted, but not definitely known that the predictability of the atmosphere decays so rapidly that a weeklong "prediction" is no better than a sheer guess. Whatever is actually the case, it is important to establish the maximum range of predictability, beyond which the *details* of the state are essentially unpredictable. In this same connection, NAMIAS (1947) and others have suggested that certain *statistical* characterizations of the state of the atmosphere might be inherently more predictable than the *details* of its state. Again, the practical question will remain unresolved until it is definitely known whether or not such statistics are in fact more predictable, and (if so) how much more predictable, and whether or not the possible gain in predictability is more than overbalanced by the attendant loss of detail or information content of the prediction.

Still another problem is that of estimating the probable error or limits of confidence in predictions, a question that bears directly on the gradual growth of error toward the level of complete "unpredictability", and one which is frequently brought up as an objection to any deterministic approach to prediction. That such objections are valid is sometimes admitted even by dynamic meteorologists (who have borne the brunt of criticism); they have not, however, been too ready to inject a jarring note of probability into what is otherwise a pleasingly deterministic theory—a course that is nevertheless inevitable if the existence of error is fully recognized.

Between the extremes of theory and practice lies the problem of detecting when the development of more and more general, complicated, and costly methods of dynamical prediction, and more and more refined computing techniques will not produce a significant increase in the overall accuracy of predictions. It is conceivable, for example, that we are already approaching the point where the total error of prediction is comparable in magnitude to the error due to uncertainties of initial state alone. If and when this is the case, it would be pointless to channel precious scientific manpower into such narrow avenues of purely technological development, or to spend additional funds for more powerful and costly computing facilities. At that point, the only hope of improving the quality of prediction will lie in reducing the uncertainty arising from the one remaining controllable source—namely, the uncertainty inherent in ignorance of the actual state of the atmosphere at any given time.

The problems outlined above are certainly not a complete catalog, but they are sufficient to indicate that a great number and variety of problems — theoretical, practical, and administrative—bear more or less directly on the question of predictability. These same problems will be reviewed later in the light of subsequent results.

Turning to more general questions of predictability, it should be noted that, if one accepts the universal applicability of physical law, one must also concede that the future state of the atmosphere is (in a certain sense) determinate and absolutely predictable. That is, the principles of conservation of momentum, mass, and energy imply that the chains of events following two *identical* states of the atmosphere are also identical, provided equal amounts of

Figure 1 (a). A hypothetically "true" initial contour pattern.

heat energy are absorbed at equivalent times and places. This is, in fact, one of the fundamental tenets of the recently developed methodology of numerical weather prediction. The distinction between this view and the pragmatic view of predictability is simply that we cannot really say when two states are identical.

For the sake of illustration, let us adopt the plausible hypothesis that the behavior of the atmosphere is actually described by the hydrodynamical (Navier-Stokes) equations, which simply express the general conservation laws in mathematical form, and which have been successfully applied in the analysis of a wide variety of hydrodynamical phenomena. Now, it is possible to integrate those equations by approximate numerical methods with a *mathematical* error that is less than any preassigned value, no matter how small. Imagine, for example, that we first reconstruct the initial state of the atmosphere by drawing isopleths (of isobaric height, say) on a very large chart for contour intervals of ten feet or one foot, or whatever the sharpness of one's pencil allows. The initial height values will be interpolated to the nearest foot or inch at a regular network (or grid) of points spaced 100 kilometers or perhaps even 10 kilometers apart. Still in imagination, we next integrate the hydrodynamical equations by the finite-difference method, starting with the interpolated initial values at the points of this very fine grid, and approximating derivatives by differences between values at neighboring points in the grid. The significant fact is that LERAY (1931) has shown that, by making the grid finer and finer, one can make the truncation errors of such an integration arbitrarily small. Moreover, by carrying more and more significant digits in computing machines of greater and greater storage capacity, it is possible to make roundoff errors as small as desired, and so to remove the one remaining type of purely mathematical error.

This is not to say, however, that solutions of the general hydrodynamical equations computed in this fashion would agree with the future state of the atmosphere to within comparably narrow limits. Even if meteorological data contained no instrument, reading, roundoff, or transmission errors, and even if the analysts' isopleths fitted the available data perfectly, a reconstruction of the initial state from data at a finite number of points will not be the same as the true state. At various points it will differ from the true state by an amount that depends on distance from the point in question to the nearest reporting station and which, on the average, depends on the distance between reporting stations.

A typical distribution of the difference between reconstructed and true isobaric height fields is shown in Fig. 1 (c). This pattern was obtained by subtracting a hypothetical "true" height field, illustrated in Fig. 1 (a), from the reconstructed height field shown in Fig. 1 (b). The latter was constructed as objectively as possible by linear and quadratic interpolation between the circled points, where the "true" height values were given. It will be noted that the distribution of the difference between

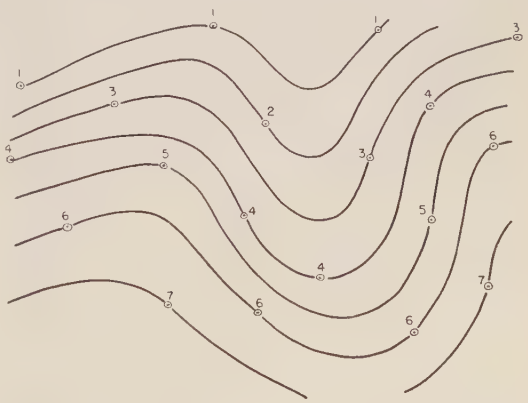


Figure 1 (b). The contour pattern reconstructed by interpolation between given "true" height values at the circled points.

Figure 1 (c). A typical pattern of initial analysis error, formed by subtracting the height field of Fig. 1 (a), from that of Fig. 1 (b). Note the relation between the "station locations" and the positions and scale of the error centers.

the true and reconstructed height fields, hereafter called the "analysis error", displays a cellular appearance. On reflection, one sees that this structure is typical, because the lines of zero error must pass through the circled points in this case and, in actuality, must pass close to the points where data are given.[1] Thus, the distribution of the analysis error has a characteristic scale or "grain-size", closely related to the distance between observing stations. As indicated earlier, the analysis error also has a characteristic magnitude, which generally increases as the distance between stations is increased.

Another important property of the analysis error stems from the fact that the large-scale transient weather disturbances are, on the average, located at random with respect to the observing stations, and vice versa. Thus, since the "phase" (e.g., zero lines) of the pattern of analysis error is essentially fixed by the locations of the stations, the analysis error tends to be uncorrelated with the true state and, in this sense, may be regarded as a random variable.

Returning to the main thread of argument, let us now suppose that the general hydrodynamical equations are integrated in the manner outlined earlier, but starting with two *different* initial states. In one case, we shall begin with the true state and, in the other, with the reconstructed state, consisting of the true state plus a field of analysis error that is random with respect to the true state. In the first case, the prediction must be correct by hypothesis. In the second, the prediction is *incorrect initially*, and will generally differ from the correct one at all later times. The amount by which it differs at various times after some specified initial instant is, of course, a measure of unpredictability or error. In fact, according to our present view, the only element of uncertainty which cannot be removed by bigger and faster computing machines is that due to the analysis error. Thus, the difference between two solutions of the hydrodynamical equations, starting with two initial states that differ by a random error field, is a measure of the *essential* unpredictability of the atmosphere. In particular, if the average magnitude of such differences at some later time approaches the average error of guessing, the atmosphere has become *essentially* unpredictable beyond that time. The remaining problem is to find out how, on the average, the unpredictability grows over an increasing period of time after an arbitrarily specified initial instant, and on what statistical properties of initial error and true state the growth rate depends.

The problem just posed, with certain idealizations to be introduced later, is substantially the question that will be subjected to mathematical analysis and examined in some detail in the remainder of this article. To state the results of this study in rather general terms, it will be found possible to express the area-averaged RMS[2] vector wind error at any time as a function of:

1) the *initial* RMS vector wind error.
2) the period of the forecast.
3) a statistical measure of the characteristic size of transient disturbances in large-scale atmospheric flow patterns.
4) a statistical measure of the scale of the "analysis error"—which, in turn, depends on the distance between adjacent observing stations.
5) the RMS vector deviation of the wind at about 500 mb from its area average.
6) the RMS vector deviation of the vertical wind shear between 250 and 750 mb from its area average.

[1] In general, the analysis error is greater in average magnitude than the errors of measurement.

[2] In what follows, "RMS" is used as an abbreviation for the "root-mean-square" value taken over a very arge area. Similarly MS stands for "mean-square".

7) the area average of the vertical wind shear between 250 and 750 mb, and
8) the average static stability of the atmosphere.

That these are important (if not the dominant) factors in the predictability of the atmosphere is consistent not only with the results of numerous studies of atmospheric instability, but with intuition and general experience.

As might be expected, the growth rate of MS error is proportional to the initial MS analysis error, and is very nearly proportional to the MS vector deviation of the wind from its average—a measure of the rate of interaction between error and true state. The growth of MS error also depends critically on the relative scales of the initial analysis error and the macroscopic fluctuations in the true flow pattern, the error increasing rapidly if the distance between adjacent observing stations is appreciably greater than the characteristic half-wavelength of the true disturbances, and even diminishing if it is appreciably less. In general, the increase in the MS vector wind error over periods of a few days is proportional to the square of the period.

The results just described apply qualitatively to both quasi-barotropic and baroclinic initial states. In cases of strongly baroclinic flow, however, the effects outlined above are considerably modified by thermodynamic processes, the growth of RMS error increasing markedly with stronger vertical wind shear (horizontal temperature gradient) and smaller static stability. For initial error fields whose scale is representative of the existing observational network over the western half of the northern hemisphere, and under conditions of only moderately strong average vertical wind shear, the MS vector wind error *arising from analysis error only* may increase by 25 percent of its initial value after one day, double its initial value after two days, and so on.

It is, indeed, found that the zonally-averaged wind field is inherently more predictable than the unaveraged wind field, at least in the case of nondivergent barotropic flow. This result stems from the fact that the percentage increase in MS error of the *average* wind after a specified period is proportional to the *fourth* power of the period, rather than the *second* power. Thus, since the base of the exponent (a nondimensional measure of time) is about the same in either case, the growth of error in the averaged wind is initially slower than in the unaveraged wind. Ultimately, this difference is due to the fact that zonal averages are not contaminated *initially* by random analysis error.

From even this bare summary, it may be seen that the results outlined above have direct bearing on several aspects of predictability. In sections III—IX, we shall discuss the main results of the mathematical analysis in more detail, and interpret them in the terms of the practical and administrative problems raised at the beginning of this introduction. As much as possible, the discussion will be kept free of sustained mathematical argument, referring only to necessary definitions and final results. The mathematical development will be relegated to three appendices. The first will deal with the predictability of nondivergent barotropic flow—an exercise that contains all the essential mathematical ingredients of an analysis of more general and complicated problems, and which is a valuable preliminary to the study of baroclinic flows. The second is merely an outline of the key points in the analysis of predictability in a simple (but fairly general) type of baroclinic flow, emphasizing the analogy to the case of barotropic flow. The last is a very brief analysis of the predictability of zonally-averaged barotropic flow.

II. Problem and General Approach

The problem that will be dealt with here falls short of the very general question posed earlier in several noteworthy, but probably not crucial respects. In the first place, the equations we shall integrate are not the hydrodynamical equations in their most general form, but the equations for a mathematical idealization or "model" of the atmosphere. This model is very similar to the familiar quasi-geostrophic model originally developed by Charney (1948), which has been tested extensively and applied to routine numerical weather prediction in the past few years, and is even more closely related to the "quasi-nondivergent" model proposed by Kuo (1956).

The virtue of the quasi-nondivergent model, aside from its evident mathematical simplicity, is that it reproduces the behavior of the very large-scale slowly-moving disturbances in the true atmosphere, without exhibiting all of the

latter's meteorologically irrelevant behavior—the existence of which would be implied by the *general* hydrodynamical equations. It has been shown (on theoretical grounds) by CHARNEY (1948) and others, and fairly well verified by numerical experiment that the approximations of this simple model exclude solutions corresponding to sound and gravity waves, but leave those corresponding to the large-scale weather disturbances essentially intact. It is true, of course, that those approximations also exclude convective instability, boundary-layer turbulence, and other small-scale phenomena that bear on the predictability of the atmosphere to some extent. It is hardly conceivable, however, that these effects could *increase* its predictability. This suggests that the estimated predictability of the quasi-nondivergent model should be taken as an upper limit on the predictability of the true atmosphere.

We shall deal, specifically, with two variants of the quasi-nondivergent model. One of these is a two-level model whose state is characterized by the streamfunctions at the 250 and 750 mb surfaces, and whose equations are easily derived by approximating vertical derivatives by centred finite-differences taken over 500 mb intervals between 0 and 500 mb, 250 and 750 mb, and 500 and 1,000 mb. The other, a degenerate form of the quasi-nondivergent model, is simply the well-known nondivergent barotropic model. Its state shall be characterized by the streamfunction at the 500 mb surface.

Let us now consider two different initial states of the model. One, the true state, shall be characterized by the streamfunction ψ_0; the other, the state reconstructed from data at a finite number of points, by the streamfunction $(\psi_0 + R_0)$. The function R_0 is the analysis error. It will be left unspecified in detail, but will be assumed random with respect to ψ_0, and will take on the statistical properties of real analysis error. For simplicity, it will also be assumed that R_0 vanishes at all times around the edges of a very large region A, with the assurance that this restriction cannot much affect average conditions over the whole of the region. We next imagine that the equations for the quasi-nondivergent model have been integrated by the usual method of extrapolating over successive short intervals of time. In one case, we start integrating from the true initial state ψ_0, to obtain the "correct" state ψ at time t. In the other case, we begin with the reconstructed initial state $(\psi_0 + R_0)$, and incorrectly predict the state $(\psi + R)$ at time t. Thus, at each point and at time t, we commit an error R.

Now, consider a statistical measure of error E, defined as the MS value of the *gradient* of R taken over the whole of the area A. Owing to the relation between wind and streamfunction, E is the MS vector wind error. We choose this particular measure of error partly because we are primarily interested in predicting the wind (gradient of streamfunction), rather than the *absolute* size of the streamfunction, and partly because the form of the equations implies that the absolute size of the streamfunction is physically irrelevant and, for that matter, unpredictable.

The question is this: How does the MS vector wind error E vary with time? On what statistical properties of the initial state ψ_0 does the growth of E depend, and in what way? How does the growth of E depend on the statistical properties of R_0—e.g., average distance between observing stations, and RMS analysis error?

The most direct approach to these problems, of course, would be to carry out comparative numerical integrations of the equations for the quasi-nondivergent model. That is, one numerical prediction made from a reconstructed initial state over a region where upper air data are actually quite dense could be compared with another made from an initial state independently reconstructed from a fraction of the actual data. It might even be sufficient to compare forecasts made from reconstructed initial states with forecasts made from the same initial states, but to which random initial error fields (with the correct *statistical* properties) have been added. There are, however, two disadvantages in this approach. The first and most obvious is that the difference between pairs of comparative forecasts would be due partially to roundoff and truncation errors, and would not isolate the *essential* unpredictability due to analysis error. The second, and probably more important disadvantage is that the error E would not be expressed as a function of the relevant statistical parameters, but as a collection of numerical values. Thus, unless one carried out an enormous number

of integrations, it would be extremely unlikely that he would discover the correct combination of factors on which the error depends. It seems desirable, therefore, to devise some analytic approach to the problem, if only as a supplement to numerical experiment.

The procedure that will be followed in the detailed mathematical analysis of the problem is this: We begin by calculating the first time-derivative of E, which is related, by definition, to the first time-derivative of the error R at each point. The latter is then expressed in terms of *space-derivatives only* by making use the equations for the quasi-nondivergent model. Similarly, we form the second time-derivative of the error E, and again eliminate the first time-derivative of R by substitution from the equations for the model. In this way, by successive differentiation and substitution from the equations for the quasi-nondivergent model, all the time-derivatives of E at any time can be expressed *in terms of the values of* ψ *and* R *current at that time.*

The next step in the basic procedure is to evaluate the time-derivatives of the error E at the initial instant, taking advantage of the fact that R *is initially random.* Owing to the latter property, the initial values of the time-derivatives of E take on relatively simple integral forms that can be expressed approximately in terms of the MS gradients of the initial streamfunction ψ_0 and analysis error R_0 and the auto-correlation functions for ψ_0 and R_0. The final step is simply to represent the error E at any time t as a Taylor series in ascending powers of t, in which the coefficients are the time-derivatives of E at the initial instant ($t = 0$). This series explicitly relates the statistical measure of error E to the statistical properties of the initial wind and error fields, as well as to the period of the forecast.

III. The Predictability of Barotropic Flow

The result of applying the procedure outlined above in the case of nondivergent barotropic flow is most concisely expressed in a formula for the percentage change in the MS vector wind error E over a given interval of time t. With certain approximations of integration, discussed more fully in Appendix I,

$$\frac{E - E_0}{E_0} = (M^2 - m^2)\,\overline{\nabla\Psi \cdot \nabla\Psi}\, t^2 F(M,m) \quad (1)$$

The function Ψ is related to the initial streamfunction ψ_0 as follows:

$$\psi_0 = \Psi - Uy$$

where U is the zonal component of the average wind, taken over the entire area A, and y is the coordinate distance toward the north. Thus, since the meridional component of the average wind must vanish, $\overline{\nabla\Psi \cdot \nabla\Psi}$ is the square of the vector deviation of the wind from its average. The bar placed above a quantity denotes its area average, taken over the entire region A. The quantity M is an inverse measure of the scale of fluctuations in the true initial flow pattern, defined as

$$M^2 = \frac{\overline{\nabla\Psi \cdot \nabla\Psi}}{\overline{\Psi^2}}$$

Similarly, m is an inverse measure of the scale of the initial error field, defined as

$$m^2 = \frac{\overline{\nabla R_0 \cdot \nabla R_0}}{\overline{R_0^2}}$$

The function $F(M, m)$ takes on either of two forms, depending on whether m exceeds M, or vice versa.

$$F(M,m) = \frac{5}{72}\begin{cases}\dfrac{M^2}{m^2} & \text{if } m > M \\ \dfrac{m^2}{M^2} & \text{if } m < M\end{cases}$$

As might be expected, the change in error $(E - E_0)$ is proportional to the initial error E_0. It is also proportional to the MS value of the vector deviation of the wind from its average, a measure of the rate at which the various interactions and exchange processes operate.

The most startling and significant implication of Eq. (1) is that the growth of error is evidently very sensitive to the difference in the scales of the initial error field and the fluctuations in the true initial flow pattern. If, for example, the scale of the initial error field were appreciably greater than that of the true fluctuations, the error would increase rapidly and the predictability of the winds would correspondingly decrease. Conversely, if the scale of the initial error field were appreciably less than that of the true fluctuations, the error might actually *decrease* for a while. On the face of it, this result seems paradoxical. It is, how-

ever, consistent with the results of independent studies of the stability of barotropic flow made by LORENZ (1953) and THOMPSON (1957), both of which indicate that the kinetic energy of perturbations decreases if their scale is less than that of meridional fluctuations in the speed of a nonuniform zonal current, and increases if it is greater. One may, in fact, interpret R_0 as a real perturbation instead of a random initial error, and regard Eq. (1) as a generalized criterion for the stability of nonequilibrium flows with respect to random perturbations of finite amplitude.

Evidence of the effectiveness of this mechanism of barotropic stability is contained in the fact that wind fields predicted by solving the barotropic vorticity equation grow noticeably "smoother" and more "zonal" as the forecast period is increased—smoother than can be accounted for by the artificial averaging process that is applied intermittently to reduce small-scale truncation error. In general, the mechanism of *baroclinic* instability (by which the actual fluctuations in the initial flow pattern were originally produced) tends to amplify components whose scale lies in the range of "barotropic damping". Thus, since the method of prediction does not distinguish between analysis error and real fluctuations of comparable or slightly greater scale, the phenomenon described above may be interpreted as a gradual transfer of kinetic energy from disturbances of small scale to those of larger scale through the mechanism of barotropic stability. Indeed, the recent work of PHILLIPS (1956) indicates that this is the principal mechanism by which the kinetic energy of the westerlies is maintained.

In summary, there is strong reason to believe that the effect of scale on the predictability of a truly barotropic atmosphere would be a very real one. Although it is peculiar to barotropic flow, it is also one of the dominant effects in strongly baroclinic flows.

One never observes, of course, that the errors of barotropic forecasts decrease as the period is increased. A part of this might be attributable to truncation error. As suggested earlier, however, a more important reason for the invariable growth of error probably lies in the fact that baroclinic flows are unstable with respect to perturbations of smaller scale than are barotropic flows—or, more obviously, the fact that the atmosphere is not really barotropic. Thus, in order to come to any definite conclusions about the predictability of the atmosphere, we must investigate the growth of error in predictions based on a baroclinic model.

Before proceeding to the discussion of more general results, it should be pointed out that Eq. (1) and all subsequent expressions for the error growth contain only the first few terms of the complete Taylor expansion, and are probably not valid for periods longer than a few days. The error cannot, for example, grow indefinitely, since its maximum value is limited by the total initial kinetic energy. Likewise, it cannot *decrease* indefinitely, for it must always remain positive. In particular, Eq. (1) is correct only to within terms involving the fourth power of the period. In that case, the first time-derivative of E vanishes at $t = 0$, as do all derivatives of odd order. The latter stems from the randomness of the *initial* error field—which, together with considerations of reversibility, implies that forecasts and "hindcasts" over periods of equal length should be equally in error.

IV. The Predictability of Baroclinic Flow

We now take up the results of applying the procedure outlined in Section II in the case of the two-level baroclinic model. With approximations that are more fully discussed in Appendix II, the percentage change in overall error during a period of length t takes the form:

$$\frac{E^\star - E_0^\star}{E_0} = t^2 \left\{ F(M,m) \left[(M^2 - m^2) \left(\frac{2m^2 + \mu^2}{m^2 + \mu^2} \right) \left(\overline{\nabla\Psi^\star \cdot \nabla\Psi^\star} + \overline{\nabla\Psi' \cdot \nabla\Psi'} \right) + \frac{m^2\mu^2}{m^2 + \mu^2} \left(\overline{\nabla\Psi' \cdot \nabla\Psi'} - \overline{\nabla\Psi^\star \cdot \nabla\Psi^\star} \right) \right] + \frac{3\mu^2}{8} \left[\frac{\mu^2 (2m^2 - M^2)}{m^2 (m^2 + \mu^2)} \overline{\nabla\Psi' \cdot \nabla\Psi'} + \frac{2\mu^2}{m^2 + \mu^2} U'^2 \right] \right\} \tag{2}$$

Here, for mathematical simplicity, the statistical measure of error $E^\star$ was taken to be

$$E^\star = \frac{1}{2} \left(\overline{\nabla R_1 \cdot \nabla R_1} + \overline{\nabla R_2 \cdot \nabla R_2} \right) + \mu^2 \left(\overline{R_1^2} + \overline{R_2^2} \right)$$

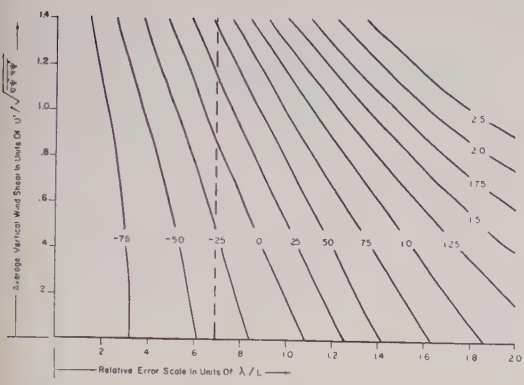


Figure 2. Percentage increase of error as a function of error scale and average vertical wind shear.

in which the subscripts 1 and 2 refer to conditions at the 250 and 750 mb surfaces, respectively, and μ^2 is a positive constant. Since $E^\star$ is still positive definite, its effectiveness as a measure of predictability is unimpaired. The quantities M, m and $F(M, m)$ are as defined in Section III, and are assumed to have the same values at 750 mb as they do at 250 mb. The constant μ, which is essentially a measure of temperature lapse-rate, is closely related to the wave number of maximum baroclinic instability, and is defined as

$$\mu^2 = 2f^2 \left(\frac{R^2 T^2}{g\Theta} \frac{\partial \Theta}{\partial z} \right)^{-1}$$

where f is the Coriolis parameter; R, the gas constant; g, the gravitational acceleration; T and Θ are representative values of the absolute temperature and potential temperature, respectively. The variables $\Psi^\star$ and Ψ' are related to the initial streamfunctions ψ_1 (at 250 mb) and ψ_2 (at 750 mb) as follows:

$$\Psi^\star = \frac{\Psi_2 + \Psi_1}{2} \qquad \Psi' = \frac{\Psi_2 - \Psi_1}{2}$$

$$\psi_1 = \Psi_1 - U_1 y \qquad \psi_2 = \Psi_2 - U_2 y$$

where U_1 and U_2 are the zonal components of the area-averaged wind at 250 and 750 mb, respectively. Thus, $\overline{\nabla\Psi^\star \cdot \nabla\Psi^\star}$ may be thought of as the MS vector deviation of the *vertically averaged* wind from its *area average*. Similarly, $\overline{\nabla\Psi' \cdot \nabla\Psi'}$ is to be interpreted as the MS vector deviation of *half* the vertical wind shear between 250 and 750 mb from its area average. Finally, U' is simply half of the zonal component of the area-averaged vertical wind shear between 250 and 750 mb.

According to Eq. (2), as in the case of barotropic flow, the growth rate of error is proportional to the initial value of the MS vector wind error. Moreover, since the term whoes coefficient is $(M^2 - m^2)$ is one of the dominant terms on the righthand side of Eq. (2), the growth rate is roughly proportional to the MS vector deviation of the vertically-averaged wind from its area average. For the same reason, it is evident that a general decay of predictability in baroclinic flow is favored by an initial error field whose scale is large relative to the scale of initial disturbances in the true flow pattern, and discouraged by the reverse.

The results described above, of course, apply equally well in the case of nondivergent barotropic flow. In the present case, however, they are considerably modified by the thermodynamic processes operative in baroclinic flow. It should be noted that the term in Eq. (2) involving U' is the only one that does not depend on deviations from horizontally uniform zonal flow and is, therefore, frequently dominant. It invariably increases the rate of error growth, by an amount that is proportional to the square of the average vertical wind shear and inversely proportional to the square of the average static stability. In other words, the atmosphere tends to be unpredictable in situations of strong horizontal temperature gradient and near-adiabatic lapse-rate—a result that is probably not very surprising to the practicing forecaster.

The joint effect of scale and vertical wind shear is illustrated in Figure 2, on which $(E^\star - E_0^\star)/E_0$ after one day is plotted as a function of average vertical wind shear and the ratio of error scale λ to the scale L of fluctuations in the true initial flow pattern. The vertical shear is expressed in units of the RMS vector deviation of the vertically averaged wind from its area average. The nondimensional numbers $t^2M^2\,\overline{\nabla\Psi^\star \cdot \nabla\Psi^\star}$ and μ^2/M^2 were assigned fixed values of 5 and 0.56, respectively—both of which are fairly representative of actual conditions in the atmosphere. The ratio of $\overline{\nabla\Psi^\star \cdot \nabla\Psi^\star}$ to $\overline{\nabla\Psi' \cdot \nabla\Psi'}$ was set equal to 4.

Figure 2 shows that the percentage growth of error in the case of no average vertical wind shear is much like that in barotropic flow, the error increasing when the scale of the initial error

field is greater than that of the true fluctuations in the initial flow pattern, and vice versa. As the vertical shear increases, however, the rate of error growth for initial error fields of any fixed scale also increases. Simultaneously, the region of scale and shear over which the growth rate is positive includes initial error fields of smaller and smaller scale. This general result is in accord with studies of baroclinic instability made by FJØRTOFT (1950), PHILLIPS (1951), EADY (1952), and THOMPSON (1953), all of whom find that increased vertical wind shear and decreased static stability tend to destabilize baroclinic flows, and also points up the close connection between the problem of predictability and the stability problem.

The seriousness of these results can be most quickly appreciated by estimating the ratio of the scale of actual fields of initial error to the scale of fluctuations in atmospheric flow patterns. Now, the average distance between regular radiosonde, rawinsonde, and dropsonde reports from land-stations, fixed ship positions, and weather reconnaissance aircraft is about 700 miles—when taken over North America, the Caribbean Sea, the North Atlantic and most of the Pacific, and weighted according to the area that each report represents.[1] Identifying the average distance between reports with the characteristic half wavelength of the initial error field, and taking the average *half* wavelength of fluctuations in the true flow pattern to be about 1,000 miles, we see that the existing circumstances lie somewhere near the dashed line on Figure 2. Thus, when the relative wind shear falls below 0.9, the *inherent* predictability of the atmosphere is quite high. (This is not to say, of course, that forecasts in such situations will be correct, for they are actually subject to truncation errors, boundary errors, and defects of the physical model.) On the other hand, there are many winter weather situations in which the relative wind shear exceeds 1.0, in which case $E^\star$ may increase by as much as $\cdot 5\ E_0$ after one day, $2.0\ E_0$ after two days, $4 \cdot 5\ E_0$ after three days, and so on. Accordingly, if the initial MS vector wind error E_0 is a substantial fraction of the MS vector wind, the error $E^\star$ rapidly approaches the level of complete unpredictability.

[1] All reports obviously cannot be given equal weight, as one can easily see in the extreme case when all reports are clustered around one point.

V. The Maximum Range of Predictability

Having gained a general idea of the rate at which initial analysis error contaminates a forecast, we are now in a position to make a crude estimate of the maximum time-range of predictability, beyond which the atmosphere is *essentially* unpredictable. Before doing so, however, one must first decide what he means by "unpredictable". Now, it is certainly in accord with the common-sense meaning of the word to regard the occurrence or non-occurrence of an event as "unpredictable", if any prediction is no better than a guess. Thus, we may shift the burden of definition to that of defining a standard method of "guessing" and a standard error of guessing, against which the errors of "predictions" can be judged. We shall then say that the atmosphere is unpredictable beyond the time when the *inherent* errors of prediction approach the error of guessing.

One of the most obvious methods of "guessing" is simply to select values of wind speed for each point at random, from a population of values whose frequency distribution is that of a representative sample. A meteorologist, of course, would not guess in this fashion, but would insure that his "guess" varied smoothly from point to point. This does not affect the error of random selection, however, for such modifications are equivalent to selecting homogeneous subpopulations from a collection of subpopulations whose frequency distribution is the same as for individual values. It is easily shown that the MS vector wind error of this type of guessing is exactly twice the MS vector deviation of the wind from its average.

On a little reflection, one sees that the procedure described above is not a good method of guessing, simply because always placing one's bet on the average wind would result in a MS vector wind error that is exactly half that of random selection. Accordingly, the RMS vector wind error of guessing will be taken to be the RMS vector deviation of the wind from its area average. It will be found convenient to express the RMS error of guessing in units of the RMS analysis error. In these units, the RMS error of guessing is of the order of 4 over most of the western half of the northern hemisphere.

Let us now suppose that $(E^\star - E_0^\star)/E_0$ after

one day is of the order of .50, a figure that is made plausible by the results of Section IV. Thus, since $\mu/^2M^2$ is approximately .5, $E^\star$ is approximately $2E$, so that

$$E \simeq E_0 \left(1 + \frac{t^2}{4}\right) \text{ for } t \text{ in days}$$

The question is now reduced to asking how big t must be in order that E (the MS vector wind error of prediction) equals the MS error of guessing—which, in units of E_0, is 16. The period τ in question is evidently

$$\tau \simeq \sqrt{60} \text{ days} \simeq 7.7 \text{ days}$$

According to this estimate, the atmosphere is *essentially* unpredictable beyond a period of about a week.

The result stated above should not be taken quite literally, since the MS vector wind error of prediction undoubtedly does not continue to increase like the square of the period beyond two or three days. Offsetting this, however, are the facts that we have considered only the *inherent* error due to uncertainty of initial state and have disregarded the various other types of error and, second, that the quasi-nondivergent model is probably more predictable than the true atmosphere.

These estimates, although they are admittedly and necessarily rather crude, must be of the correct order of magnitude and are consistent with general experience. Verification studies carried out by Brier (1944) and others are in unanimous agreement that the forecaster loses virtually *all* ability to predict the details of a weather situation for periods longer than about five days. The significance of the present results does not, however, lie in merely explaining the *observed* decay of accuracy, but in being able to predict what the decay of predictability would be in circumstances that are altogether different from the present one —e.g., if the data density were drastically increased or reduced.

Another point deserving emphasis is that the foregoing results apply to the predictability of flow patterns *in natural detail*, or to meteorological events that depend on the phases of individual disturbances—as, for instance, rain or overcast skies on a particular day. These results do not necessarily apply to the predictability of space- or time-averaged flow patterns. As mentioned earlier, it has been suggested by Namias (1947) that averaged wind fields might be inherently more predictable than detailed flow patterns, and might be predicted over longer periods. In the next section, we shall discuss the predictability of zonally-averaged flow patterns in a case where direct comparisons are possible, namely, in nondivergent barotropic flow.

VI. The Predictability of Zonally-Averaged Flow

In an earlier paper, Thompson (1957) has developed a method for predicting the zonally-averaged flow of a nondivergent barotropic fluid. This enables us to compare the predictability of zonally-averaged flow with that of unaveraged flow in a case that resembles the actual situation enough to warrant a few tentative conclusions. Making use of techniques similar to those outlined in Section II, and introducing approximations discussed more fully in Appendix III, we find that

$$E \simeq \left[\frac{1}{4}\left(\overline{M}^2 - 2m^2\right)E_0 t^2\right]^2 \frac{1}{W}\int_0^W U^2 dy \quad (3)$$

in which m, E_0, and t are as defined in Section III, and W is the entire width of the flow. The quantity U is the difference between the zonally-averaged eastward component of the wind at a particular latitude and its average taken over all latitudes; E is now the MS error in U taken over all latitudes. The constant $\overline{M}$ is the "characteristic wave number" of meridional fluctuations in U.

According to Eq. (3), the statistical measure of error (when normalized with respect to variations in the quantity to be predicted) is equal to the *fourth* power of a nondimensional measure of the forecast period. Referring back to Eq. (1), we see that the corresponding normalized error in predictions of the *unaveraged* flow can be put in a very similar form, the non-dimensional measure of the forecast period being about the same, but raised to the second power. This implies that the zonally-averaged flow is more predictable than the unaveraged flow for periods during which the nondimensional measure of time is less than unity. Superficially, at least, it also implies that the zonally-averaged flow becomes *less* predictable than the unaveraged flow over longer

periods, a conclusion that should not be taken seriously in view of the obvious limitations of truncated Taylor expansions over finite intervals. It is hardly conceivable that a quantity which is more predictable over short periods should be less predictable over long periods.

The greater predictability of zonally-averaged flow—which stems directly from the fact that the second order time-derivative of the error vanishes initially—is due to an essential difference between the initial states of averaged and unaveraged flows. The reason is simply that the averaged flow is not contaminated initially by random analysis error. Although it is certainly not conclusive, this result suggests that any average which is not sensitive to analysis error is more predictable than the unaveraged flow pattern. At the same time, it must be pointed out that the possible increase of predictability brought about by averaging is acquired at the price of losing "information content"—in the usual sense that the prediction of averages tells less about the unexpected or abnormal event.

VII. Predictability as a Measure of Confidence in Forecasts

Another question that was raised in the introduction and which needs little elaboration is that of estimating the probable error of predictions. Although Eq. (2) says nothing about the total error to be expected from approximations of all types, it does provide a means of calculating the inherent error due to the uncertainty of initial state, and a means of rating forecasts according to the degree of confidence we should place in them. In general, of course, the probable error does not depend only on the initial error field, whose statistics are fixed by the character of the observational network, but also depends on conditions of scale, vertical wind shear, and static stability in the particular meteorological situation in question. Thus, the inherent error should (and can) be computed daily, and issued in conjunction with the forecast itself.

By this time, it should be abundantly clear that the theory described here is a combination of the dynamical and probabilistic approaches to prediction. The probabilistic element is introduced directly into the dynamical equations by specifying only the statistics of the analysis error; this, in fact, is all that is known about an essentially random error field. It is not claimed that this is the only approach, or even the best approach. It is merely a first step toward reconciling a deterministic approach with the meteorological facts of life.

VIII. The Point of Rapidly Diminishing Returns in the Development of Numerical Forecasting Methods

Our next concern is to compare the total error of predictions with the inherent error due to uncertainty of initial state alone. Experience over the past few years indicates that the RMS vector wind error of numerical predictions over 36 hours is about 50 % of the RMS vector deviation of the wind from its area average. Judging from the rough estimates made in Section V, the *inherent* RMS error after 36 hours is about 30 % of the same standard. The residual error—roughly 40 % of the RMS vector deviation of the wind from its average—must be due to the combined effects of truncation error, round-off error, boundary error, and shortcomings of the physical model. A part of this error, conceivably, might also be due to some fundamental indeterminacy in the behavior of the atmosphere.

The result stated above is subject to two rather important interpretations. In the first place, errors of numerical method and defects of the present physical models are quite large enough to account for the residual error *not accounted for by the inherent error*, and are controllable within economic limitations. In short, there is nothing in these results to indicate that the atmosphere is unpredictable in any fundamental sense—i.e., that there is any mysterious principle of uncertainty operating, other than the very real one discussed at length in this article.

It is also interesting to note that the inherent RMS error after 36 hours is about the same size as the residual RMS error, both being about one third of the RMS deviation of the wind from its average. This indicates that it is certainly worth trying to halve the residual error through the use of higher order finite-difference approximations, special smoothing techniques, hemispheric finite-difference grids, more reference levels, and more realistic

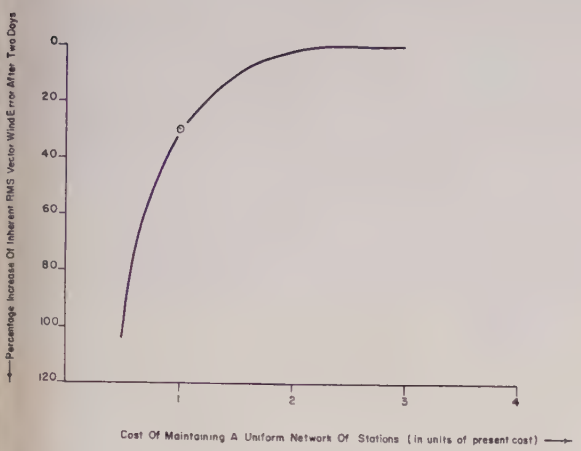


Figure 3. The cost of maintaining a uniformly dense network of stations, plotted against the percentage increase of inherent error corresponding to that outlay. The circled point on the curve represents the existing situation.

physical models. It is also evident that even further efforts in this direction would at most result in something like a 15 % reduction in a total RMS error that would still be about 35 % of the RMS deviation of the wind from its average, and that such further gains in accuracy would be bought at enormous cost. In summary, the development of better and more complicated numerical prediction methods has not yet reached the point of rapidly diminishing returns, but will have approached it when the RMS error due to "controllable" sources has been about halved. At that point, the only significant gain in actual predictability is to be made by reducing the inherent error—that is, by increasing the density of regular reporting stations. It is also clear that the time to prepare for such drastic action has already come.

IX. The Effect of Increasing the Density of Observing Stations.

It has been stressed several times earlier that the growth of the inherent error—due to errors in reconstructing the initial state from data at a finite number of points—depends crucially on the scale of the initial error field and, ultimately, on the spacing of the upper air observing stations. We have shown, further, that initial analysis error over the oceans and other regions of poor data coverage leads to a widespread and rapid growth of error—so rapid that the detailed flow of the atmosphere becomes completely and essentially unpredictable after periods of about a week. This situation, as has long been recognized, can be remedied effectively only by increasing the density of reporting stations over regions where the density is now quite low. The remaining questions are these: What return in increased predictability can be expected from increasing the overall density of reporting stations, and how does this compare with the corresponding outlay of funds? Where is the point of rapidly diminishing return per outlay? How should new stations be located in effecting an increase of overall station density?

The facts bearing on these questions are summarized in Figure 3, which shows a graph of the cost of maintaining a uniformly dense network of stations (expressed in units of the cost of maintaining a uniformly dense network, whose density is equivalent to the overall density of the present network) plotted against the percentage increase of the *inherent* RMS vector wind error after two days for a relative shear of 1.2. For simplicity, it was assumed that the cost of maintenance per observation is the same for all stations, fixed or moving, regardless of type. This assumption would be fairly true, for example, of land stations and installations on ships moving along their normal courses. Our present situation is estimated to lie near the circled point on the curve.

Inspecting Figure 3, we see that doubling the cost and overall density of observing stations would all but eliminate the increase of *inherent* error. Thus, although it would not completely eliminate the initial error, it would reduce the *total* RMS vector wind error in 48 hour predictions to something on the order of 50% of the RMS deviation of the wind from its average, rather than the existing error of about 60 %. Moreover, according to our present estimates, trebling the cost and overall density of the observational network would produce no significant reduction of inherent error, beyond that attainable by doubling the present outlay. Thus, the point of rapidly diminishing returns lies somewhere around double the present cost.

In actuality, of course, the density of observing stations is far from uniform. Thus, the optimum strategy in effecting an overall increase in density is to establish stations in the

regions where the density is lowest up to the density level of the surrounding regions, and then to increase the density over a more inclusive region up to the density level of areas of good coverage. This is simply because the cost of producing an *effective* increase of *overall* density in a *nonuniform* network is less than the cost of producing the same increase of density in a *uniform* network. In fact, if economy were a strong factor, and if the aims were to produce uniform forecasting accuracy over very large areas, it would be desirable to replace some land stations in regions where they are extremely dense by stations aboard moving ships.

From considerations of symmetry alone, the optimum distribution of observing stations is very nearly one of uniform density, slightly modified by the varying degree of meteorological activity over different regions and by the fact that maximum forecasting accuracy should be attained in regions of dense population. Experience with numerical prediction over the past two years has shown that two and three day forecasts for the United States alone are strongly dependent on the initial state of the atmosphere over more than half a hemisphere. Accordingly, the fact that few people live or travel in the mid-Pacific is not a strong argument for concentrating observing stations in the United States.

It is hoped that the results described in the foregoing sections of this article, or more refined estimates based on a similar theory of predictability, can be used as a common basis for decision and action in dealing with a problem that, by its very nature, is truly international. It is also to be hoped that these rather crude estimates of predictability will not be accepted at face value, but will provoke more complete studies of the empirical factors that enter into this important problem.

X. The Effects of Heat Sources and Viscosity

The estimates of predictability presented earlier do not include the effects of viscous momentum transfer and nonadiabatic heating. Thus, although these processes are probably not important factors in the predictability of atmospheric flow patterns over short periods of time, it is still a matter of interest to speculate about their general effect. By and large, the net effect of internal viscosity is to reduce the variance of relative vorticity, and to obliterate gradually any trace of previously existing circulation centers. Consequently, the state of a viscous fluid depends more on the dynamical properties of the system and the external influences acting on it than it does on an initial state at some time in the remote past. It is hard to imagine, therefore, that the effect of viscosity is to *increase* the predictability of the flow, except in a sense to be discussed later.

Similarly, it is difficult to conceive that nonadiabatic heating acts in such a way as to increase the predictability of the *detailed* flow pattern over short periods. Apart from the fact that the distribution of heat sources is not precisely known, it is improbable that the existence of convective instability on any scale is conducive to accurate prediction. It is also true, of course, that nonadiabatic heating has a decisive effect on long period trends in (say) monthly averages—as, for instance, the change of seasons. It is certainly not necessary to know the detailed state of the atmosphere in the middle of the summer in order to predict the coming of winter. This is due to the fact that long-period averages depend very little on the detailed state of the atmosphere at some time in the far distant past (partly through the dissipative action of viscosity), but are positively controlled by the cumulative effect of energy input. It may turn out, in fact, that averages whose variations depend primarily on the external influences exerted on a quasi-linear system are more predictable over monthlong periods than the detailed state is over a week or so.

XI. Summary of Results

A brief review of the main results is given in the last five paragraphs of the introduction.

Appendix I—The Predictability of Nondivergent Barotropic Flow

We begin with the definition of the mean square error E.

$$E = \frac{1}{A}\int_A \nabla R \cdot \nabla R \, dA$$

where A is the area of a very large region (also designated by A), dA is an element of A, and R is the error in the streamfunction at various points and at time t. All notation is standard,

unless specified otherwise. Differentiating E with respect to time, we have

$$\frac{\partial E}{\partial t} = \frac{2}{A}\int_A \nabla R \cdot \nabla \frac{\partial R}{\partial t}\, dA$$

$$= \frac{2}{A}\int_A \left(\nabla \cdot R \nabla \frac{\partial R}{\partial t} - R\nabla^2 \frac{\partial R}{\partial t} \right) dA$$

A part of the integral above may be transformed by applying Gauss' theorem, as follows:

$$\frac{\partial E}{\partial t} = \frac{2}{A}\oint_C R \frac{\partial}{\partial n}\left(\frac{\partial R}{\partial t}\right) dC - \frac{2}{A}\int_A R\nabla^2 \frac{\partial R}{\partial t}\, dA$$

in which C is the path around the boundary of A, dC is an element of C, and a derivative with respect to n is the component of the vector gradient normal (outward) to C. Thus, since R vanishes on C by hypothesis,

$$\frac{\partial E}{\partial t} = -\frac{2}{A}\int_A R\nabla^2 \frac{\partial R}{\partial t}\, dA \qquad (4)$$

Now, the equation governing the evolution of the flow pattern from an initially *incorrect* wind field is

$$\nabla^2 \frac{\partial}{\partial t}(\psi + R) + J(\psi + R,\ \nabla^2\psi + \nabla^2 R) + \beta \frac{\partial}{\partial x}(\psi + R) = 0$$

whereas the equation for the *correct* flow pattern is

$$\nabla^2 \frac{\partial \psi}{\partial t} + J(\psi,\ \nabla^2\psi) + \beta \frac{\partial \psi}{\partial x} = 0 \qquad (5)$$

β is, as usual, the Rossby parameter and x is the eastward coordinate. Subtracting the equation above from the one preceding it,

$$\nabla^2 \frac{\partial R}{\partial t} = -\left[J(R, \nabla^2\psi + \nabla^2 R) + J(\psi, \nabla^2 R) + \beta \frac{\partial R}{\partial x} \right] \equiv -N \qquad (6)$$

We next eliminate $\nabla^2 \frac{\partial R}{\partial t}$ between Eqs. (4) and (6), with the result that

$$\frac{\partial E}{\partial t} = \frac{2}{A}\int_A RJ(\psi,\ \nabla^2 R)\, dA + \frac{1}{A}\int_A J(R^2,\ \nabla^2\psi + \nabla^2 R)\, dA + \frac{1}{A}\int_A \beta \frac{\partial R^2}{\partial x}\, dA \qquad (7)$$

By applying Stokes theorem, the second integral on the right-hand side of Eq. (7) can be transformed into a line integral around C, whose integrand contains R as a factor; thus, since R vanishes on C, that integral vanishes. For similar reasons, the third integral on the right-hand side of Eq. (7) vanishes. At any time t, therefore,

$$\frac{\partial E}{\partial t} = \frac{2}{A}\int_A RJ(\psi,\ \nabla^2 R)\, dA \equiv \frac{2}{A}\int_A \mathbf{k} \cdot \nabla\psi \times R\nabla(\nabla^2 R)\, dA \qquad (8)$$

where $\mathbf{k}$ is a unit vector, directed vertically upward. At $t = 0$, ψ and R are uncorrelated by hypothesis, so that the right-hand side of Eq. (8) vanishes *initially*—i.e., designating conditions at $t = 0$ by the subscript zero,

$$\left(\frac{\partial E}{\partial t}\right)_0 = 0 \qquad (9)$$

We now differentiate Eq. (8) with respect to time. At any time,

$$\frac{\partial^2 E}{\partial t^2} = \frac{2}{A}\int_A \mathbf{k} \cdot \nabla\psi \times \left[\frac{\partial R}{\partial t}\nabla(\nabla^2 R) + R\nabla\left(\nabla^2 \frac{\partial R}{\partial t}\right) \right] dA + \frac{2}{A}\int_A \mathbf{k} \cdot \nabla\left(\frac{\partial \psi}{\partial t}\right) \times R\nabla(\nabla^2 R)\, dA \qquad (10)$$

At $t = 0$, however, the second integral on the right-hand side of Eq. (10) vanishes, since $\frac{\partial \psi}{\partial t}$ is given in terms of ψ *alone* by Eq. (5), and because ψ and R are uncorrelated *initially*. Initially, therefore,

$$\frac{\partial^2 E}{\partial t^2} = \frac{2}{A}\int_A \mathbf{k}\cdot\nabla\psi\times$$

$$\left[\frac{\partial R}{\partial t}\nabla(\nabla^2 R) - \left(\nabla^2\frac{\partial R}{\partial t}\right)\nabla R\right] dA$$

$$+\frac{2}{A}\int_A J\left(\psi,\ R\nabla^2\frac{\partial R}{\partial t}\right) dA \qquad (11)$$

By applying Stokes' theorem, the second integral on the right-hand side of Eq. (11) can be transformed into a line integral around C, whose integrand contains a factor of R; thus, since R vanishes on C, that integral also vanishes.

The next step is to eliminate the time derivatives of R from Eq. (11) by resubstitution from Eq. (6). We have

$$\nabla^2\frac{\partial R}{\partial t} = -N,$$

from which

$$\frac{\partial R}{\partial t} = \iint_A G(\xi,\ \eta)\, N(x+\xi,\ y+\eta)\, d\xi d\eta \qquad (12)$$

where G is the Green's function corresponding to the Laplace operator, for $G = 0$ on C; x and y are rectangular coordinates in a horizontal plane; ξ and η are variables of integration corresponding to x and y. As implicitly assumed in Eq. (12), the Green's function for a large region A is very nearly independent of x and y at almost all points of the region. Similarly, letting $\nabla^2 R = Z$,

$$\nabla^2(\nabla R) \equiv \nabla(\nabla^2 R) = \nabla Z$$

$$\nabla R = -\iint_A G(\xi,\ \eta)\,\nabla Z(x+\xi,\ y+\eta)\, d\xi d\eta$$

Thus, substituting these results into Eq. (11),

$$\frac{\partial^2 E}{\partial t^2} = \frac{2}{A}\iint_A \mathbf{k}\cdot\nabla\psi\,(x,\ y)\times\left[\nabla Z(x,\ y)\right.$$

$$\iint_A G(\xi,\ \eta)\ N(x+\xi,\ y+\eta)\ d\xi d\eta - N(x,\ y)$$

$$\left.\iint_A G(\xi,\ \eta)\,\nabla Z(x+\xi,\ y+\eta)\, d\xi d\eta\right] dxdy$$

From now on, it is to be understood that this equation applies only at $t = 0$, and that the right-hand side is to be evaluated from the *initial* fields of ψ and R. We now invert the order of integration in the equation above, with the result that

$$\frac{\partial^2 E}{\partial t^2} = 2\iint_A G(\xi,\ \eta)\,\chi\,(\xi,\ \eta)\, d\xi d\eta \qquad (13)$$

where

$$\chi(\xi,\ \eta) = \frac{1}{A}\iint_A \mathbf{k}\cdot\nabla\psi(x,\ y)\times[N(x+\xi,\ y+\eta)$$

$$\nabla Z(x,\ y) - N(x,\ y)\,\nabla Z(x+\xi,\ y+\eta)]\, dxdy$$

Finally, introducing the definitions of N and Z,

$$\chi(\xi,\ \eta) = \frac{1}{A}\iint_A\left\{J(\psi,\ \nabla^2 R)\left[J_d(R,\ \nabla^2\psi + \nabla^2 R)\right.\right.$$

$$\left. + J_d(\psi,\ \nabla^2 R) + \beta\frac{\partial R_d}{\partial x}\right] - J(\psi,\ \nabla^2 R_d)$$

$$\left.\left[J(R,\ \nabla^2\psi + \nabla^2 R) + J(\psi,\ \nabla^2 R) + \beta\frac{\partial R}{\partial x}\right]\right\} dxdy$$

where quantities without subscript apply at the variable point (x, y), and those with subscript d apply at the point $(x+\xi,\ y+\eta)$.

Owing to the randomness of R and the fact that ψ and R are initially uncorrelated, certain terms in the integrand above vanish in the mean. Others, however, do not vanish. In fact, the integral of one such term

$$\frac{1}{A}\iint_A J(\psi,\ \nabla^2 R)\, J_d(\psi,\ \nabla^2 R)\ dxdy$$

is the autocorrelation function for the advection of initial vorticity error with the true initial wind. Similarly, $J(\psi,\ \nabla^2 R)$ is correlated with $J_d\ (R,\ \nabla^2\psi)$, simply because R (which oscillates around zero) is highly correlated with $\nabla^2 R$, and ψ is correlated with $\nabla^2\psi$. As a result of such considerations, χ may be written as

$$\chi(\xi,\ \eta) = \frac{1}{A}\iint_A [J(\Psi,\ \nabla^2 R)\, J_d(R,\ \nabla^2\Psi)$$

$$+ J(\Psi,\ \nabla^2 R)\, J_d(\Psi,\ \nabla^2 R)$$

$$- J(\Psi,\ \nabla^2 R_d)\, J(R,\ \nabla^2\Psi)$$

$$- J(\Psi,\ \nabla^2 R_d)\, J(\Psi,\ \nabla^2 R)]\, dxdy \qquad (14)$$

in which $\psi_0 = \Psi - Uy$, U is the zonal component of the average wind over A, and y is the northward coordinate. Since both Ψ and R oscillate around zero, they are both highly

correlated with their Laplacians. For purposes of approximating χ, therefore, we write

$$\nabla^2 R = -m^2 R \qquad \nabla^2 \Psi = -M^2 \Psi$$

where m and M are inverse measures of scale, defined as

$$m^2 = -\frac{\overline{R\nabla^2 R}}{\overline{R^2}} \qquad M^2 = -\frac{\overline{\Psi \nabla^2 \Psi}}{\overline{\Psi^2}}$$

and the bar above a quantity denotes its area average over A. Introducing these approximations into Eq. (14), we find that

$$\chi(\xi, \eta) \simeq m^2 (M^2 - m^2)$$
$$\left[\overline{J(\Psi, R) J(\Psi, R_d)} - \overline{J(\Psi, R) J_d(\Psi, R)}\right] \quad (15)$$

This equation makes it clear that χ is a kind of autocorrelation function.

The next problem is to express χ in terms of the autocorrelation functions for the initial fields of Ψ and R. Let us first consider the quantity

$$\overline{J(\Psi, R) J(\Psi, R_d)}$$
$$= \overline{|\nabla \Psi|\ |\nabla \Psi|\ |\nabla R|\ |\nabla R_d|\ \sin\alpha\ \sin\alpha_d}$$

in which α is the angle between $\nabla\Psi(x, y)$ and $\nabla R(x, y)$, and α_d is the angle between $\nabla\Psi \cdot (x, y)$ and $\nabla R(x+\xi, y+\eta)$. Now, it is clear that the decrease of χ, as ξ and η become larger and larger, is not due to any systematic decrease in the correlation between $|\nabla R|$ and $|\nabla R_d|$, but to a decrease in the correlation between α and α_d. Approximately, then

$$\overline{J(\Psi, R) J(\Psi, R_d)}$$
$$= \overline{\nabla\Psi \cdot \nabla\Psi}\ \ \overline{\nabla R \cdot \nabla R}\ \ \overline{\sin\alpha\ \sin\alpha_d}$$

Moreover, assuming that all angles α are equally probable in the mean,

$$\overline{\sin\alpha\ \sin\alpha_d} \simeq \frac{1}{2}\bar{\chi}_R$$

where $\bar{\chi}_R$ is the normalized autocorrelation function for the initial field of R. Following a similar line of reasoning, we find that

$$\overline{J(\Psi, R) J_d(\Psi, R)} \simeq \frac{1}{2}\overline{\nabla\Psi \cdot \nabla\Psi}\ \overline{\nabla R \cdot \nabla R}\ \bar{\chi}_R\ \bar{\chi}_\Psi$$

in which $\bar{\chi}_\Psi$ is the normalized autocorrelation function for the initial Ψ field. Introducing these estimates into Eq. (15), and substituting the resulting expression for $\chi(\xi, \eta)$ into Eq. (13), we have

$$\frac{\partial^2 E}{\partial t^2} = m^2 (M^2 - m^2)\ \overline{\nabla\Psi \cdot \nabla\Psi}\ \overline{\nabla R \cdot \nabla R}$$
$$\iint_A G(\xi, \eta)\,(\bar{\chi}_R - \bar{\chi}_R \bar{\chi}_\Psi)\, d\xi d\eta \quad (16)$$

This formula expresses the second time-derivative of E at $t = 0$ in terms of statistical properties of the initial fields of Ψ and R—namely, the characteristic scales of those fields, their mean-square gradients, and their autocorrelation functions. The Green's function G is analytic, and can be computed with any desired degree of accuracy.

The most direct procedure, of course, would be to determine the autocorrelations for Ψ and R empirically. It is desirable, on the other hand, to express $\bar{\chi}_\Psi$ and $\bar{\chi}_R$ explicitly in terms of the scale parameters m and M. Accordingly, for purposes of estimate, we shall assign them simple analytic forms. Provided the statistics of the Ψ and R fields are isotropic, the second derivatives of the autocorrelation functions at $(\xi, \eta) = (0, 0)$ are

$$\frac{d^2\bar{\chi}_\Psi}{dr^2} = -\frac{M^2}{2} \qquad \frac{d^2\bar{\chi}_R}{dr^2} = -\frac{m^2}{2}$$

where $r^2 = \xi^2 + \eta^2$. We now assume that the equations above hold from $r = 0$ to the points where the autocorrelation functions vanish. That is,

$$\bar{\chi}_\Psi = \begin{cases} 1 - \dfrac{M^2 r^2}{4} & \text{when } r < \dfrac{2}{M} \\ 0 & \text{when } r > \dfrac{2}{M} \end{cases}$$

$$\bar{\chi}_R = \begin{cases} 1 - \dfrac{m^2 r^2}{4} & \text{when } r < \dfrac{2}{m} \\ 0 & \text{when } r > \dfrac{2}{m} \end{cases}$$

To a degree consistent with earlier approximations of G, these expressions lead to:

$$\iint_A G(\xi, \eta)\,(\bar{\chi}_R - \bar{\chi}_R \bar{\chi}_\Psi)\, d\xi d\eta$$
$$\simeq \frac{5}{36} \begin{cases} \dfrac{M^2}{m^4} & \text{if } m > M \\ \dfrac{1}{M^2} & \text{if } m < M \end{cases}$$

Substitution of this result in Eq. (16) gives

$$\left(\frac{\partial^2 E}{\partial t^2}\right)_0 = 2(M^2 - m^2)\,\overline{\nabla\Psi_0 \cdot \nabla\Psi_0}\;\overline{\nabla R_0 \cdot \nabla R_0}\,F(m, M)$$

where

$$F(m, M) = \frac{5}{72}\begin{cases} \dfrac{M^2}{m^2} & \text{if } m > M \\ \dfrac{m^2}{M^2} & \text{if } m < M \end{cases}$$

Finally, expanding E in a Taylor series around $t = 0$, and recalling that $(\partial E/\partial t)_0$ vanishes,

$$E - E_0 = \frac{t^2}{2!}\left(\frac{\partial^2 E}{\partial t^2}\right)_0 + \frac{t^3}{3!}\left(\frac{\partial^3 E}{\partial t^3}\right) + \cdots$$
$$\simeq E_0\,(M^2 - m^2)\,\overline{\nabla\Psi_0 \cdot \nabla\Psi_0}\;t^2\,F(m, M)$$

From considerations of reversibility, the third derivative of E at $t = 0$ (and all higher derivatives of odd order) must vanish. The equation above is Eq. (1) of Section III. Together with Eq. (8), the result above implies that the error cannot remain random with respect to Ψ; although *initially* random, the error field is organized through nonlinear interactions to produce an increase or decrease of error as the period t is increased.

Appendix II—The Predictability of Baroclinic Flow

The procedure for calculating the initial time-derivatives of mean-square error in the case of the two-level quasi-nondivergent model is exactly analogous to the pattern of development described in some detail in Appendix I, and will be reproduced only in outline. As before, we begin by considering a statistical measure of error E.

$$E = \frac{1}{A}\int_A \frac{1}{2}(\nabla R_1 \cdot \nabla R_1 + \nabla R_2 \cdot \nabla R_2)\,dA$$
$$\equiv \frac{1}{A}\int_A (\nabla R^\star \cdot \nabla R^\star + \nabla R' \cdot \nabla R')\,dA$$

in which R_1 and R_2 are the errors in the streamfunctions at 250 and 750 mb, respectively, and

$$R^\star = \frac{R_2 + R_1}{2} \qquad R' = \frac{R_2 - R_1}{2}$$

Unless otherwise indicated, all "starred" and "primed" variables shall bear this same relationship to the corresponding variables at 250 and 750 mb. The equation analogous to Eq. (4) is

$$\frac{\partial E}{\partial t} = -\frac{2}{A}\int_A \left(R^\star \nabla^2 \frac{\partial R^\star}{\partial t} + R' \nabla^2 \frac{\partial R'}{\partial t}\right) dA \quad (4\,a)$$

In this case, the equations by which the time-derivatives of $R^\star$ and R' are eliminated from $\partial E/\partial t$ and higher derivatives of E are

$$\nabla^2 \frac{\partial R^\star}{\partial t} = -\Big[J(R^\star,\ \nabla^2\psi^\star + \nabla^2 R^\star) + J(\psi^\star,\ \nabla^2 R^\star) + J(R',\ \nabla^2\psi' + \nabla^2 R') + J(\psi',\ \nabla^2 R') + \beta\frac{\partial R^\star}{\partial x}\Big] \equiv -N^\star \quad (6\,a)$$

$$\nabla^2 \frac{\partial R'}{\partial t} = -\Big[J(R^\star,\ \nabla^2\psi' + \nabla^2 R') + J(\psi^\star, \nabla^2 R') + J(R',\ \nabla^2\psi^\star + \nabla^2 R^\star) + J(\psi',\ \nabla^2 R^\star) + \beta\frac{\partial R'}{\partial x} - \mu^2\frac{\partial R'}{\partial t} - \mu^2 J(\psi^\star + R^\star,\ R') - \mu^2 J(R^\star,\ \psi')\Big] \equiv -N' \quad (6\,b)$$

Substituting from Eqs. (6 a) and (6 b) into Eq. (4a), and introducing the fact that $R^\star$ and R' are assumed to vanish on C,

$$\frac{\partial E^\star}{\partial t} = \frac{2}{A}\int_A [R^\star J(\psi^\star,\ Z^\star) + R^\star J(\psi',\ Z') + R' J(\psi^\star,\ Z') + R' J(\psi',\ Z^\star) - \mu^2 R' J(R^\star,\ \psi')]\,dA \quad (8\,a)$$

where

$$E^\star = \frac{1}{A}\int_A (\nabla R^\star \cdot \nabla R^\star + \nabla R' \cdot \nabla R' + \mu^2 R'^2)\,dA \text{ and}$$

$$Z^\star = \nabla^2 R^\star \qquad Z' = \nabla^2 R'$$

As before, the initial randomness of $R^\star$ and R' requires that $\partial E^\star/\partial t$ vanish at $t = 0$. Differentiating Eq. (8 a) with respect to time, and again taking advantage of the facts that $R^\star$ and R' are initially random and vanish on C, we obtain an equation analogous to Eq. (11).

$$\frac{\partial^2 E^\star}{\partial t^2} = \frac{2}{A}\int_A \Big\{ \mathbf{k}\cdot\nabla\psi^\star \times \Big[\frac{\partial R^\star}{\partial t}\nabla Z^\star - \frac{\partial Z^\star}{\partial t}\nabla R^\star + \frac{\partial R'}{\partial t}\nabla Z' - \frac{\partial Z'}{\partial t}\nabla R'\Big] + \mathbf{k}\cdot\nabla\psi' \times \Big[\frac{\partial R^\star}{\partial t}\nabla Z' - \frac{\partial Z'}{\partial t}\nabla R^\star + \frac{\partial R'}{\partial t}\nabla Z^\star - \frac{\partial Z^\star}{\partial t}\nabla R'\Big] - \mu^2\frac{\partial R'}{\partial t}J(R^\star, \psi') + \mu^2\frac{\partial R^\star}{\partial t}J(R', \psi')\Big\} dA \qquad (11\,a)$$

Time-derivatives of $R^\star$ and R' are again eliminated by recourse to Eqs. (6 a) and (6 b), with the result that

$$\frac{\partial^2 E^\star}{\partial t^2} = 2\iint_A G(\xi, \eta)\,\chi(\xi, \eta)\,d\xi d\eta \qquad (13\,a)$$

where now

$$\chi(\xi, \eta) = \frac{1}{A}\iint_A [\mathbf{k}\cdot\nabla\psi^\star \times (N_d^\star \nabla Z^\star - N^\star \nabla Z_d^\star + N_d' \nabla Z' - N' \nabla Z_d') + \mathbf{k}\cdot\nabla\psi' \times (N_d^\star \nabla Z' - N' \nabla Z_d^\star + N_d' \nabla Z^\star - N^\star \nabla Z_d') - \mu^2 N_d' J(R^\star, \psi') + \mu^2 N_d^\star J(R', \psi')]\, dx dy$$

We next introduce the definitions of $N^\star$, N', $Z^\star$ and Z', and make further use of the initial randomness of $R^\star$ and R' to put χ in a form similar to that of Eq. (15).

$$\chi(\xi, \eta) = m^2(M^2 - m^2)\left(\frac{2m^2 + \mu^2}{m^2 + \mu^2}\right) \overline{[J(\Psi^\star, R_d)J(\Psi^\star, R) + J(\Psi', R_d)J(\Psi', R) - J(\Psi^\star, R)J_d(\Psi^\star, R) - J(\Psi', R)J_d(\Psi', R)]} + \frac{m^4\mu^2}{m^2 + \mu^2} \overline{[J(\Psi', R_d)J(\Psi', R) - J(\Psi^\star, R_d)J(\Psi^\star, R) - J(\Psi', R)J_d(\Psi', R) + J(\Psi^\star, R)J_d(\Psi^\star, R)]} + \frac{\mu^4(2m^2 - M^2)}{m^2 + \mu^2}\overline{J(\Psi', R)}\,\overline{J_d(\Psi', R)} + \frac{\mu^4 m^2}{m^2 + \mu^2} U'^2\,\overline{\nabla R\cdot\nabla R_d} \qquad (15\,a)$$

in which $\psi^\star = \Psi^\star - U^\star y$ and $\psi' = \Psi' - U' y$; $U^\star$ is the zonal component of the area average of the vertically-averaged wind, and U' is the zonal component of the area average of half the vertical wind shear between 250 and 750 mb. For simplicity (although it is not necessary) it has been assumed that the statistics of $R^\star$ and R' are the same. Following procedures similar to those by which Eq. (16) was derived from Eq. (15), we arrive at a formula that expresses the initial value of the second time-derivative of $E^\star$ in terms of the statistical properties of the initial error and wind fields:

$$\left(\frac{\partial^2 E^\star}{\partial t^2}\right)_0 = \overline{\nabla R\cdot\nabla R}\Big\{\Big[m^2(M^2 - m^2)\left(\frac{2m^2 + \mu^2}{m^2 + \mu^2}\right)\left(\overline{\nabla\Psi^\star\cdot\nabla\Psi^\star} + \overline{\nabla\Psi'\cdot\nabla\Psi'}\right) + \frac{m^4\mu^2}{m^2 + \mu^2}\left(\overline{\nabla\Psi'\cdot\nabla\Psi'} - \overline{\nabla\Psi^\star\cdot\nabla\Psi^\star}\right)\Big] \iint_A G(\xi, \eta)\,(\bar{\chi}_R - \bar{\chi}_R\bar{\chi}_\Psi)\,d\xi d\eta + \frac{\mu^4(2m^2 - M^2)}{m^2 + \mu^2}\overline{\nabla\Psi'\cdot\nabla\Psi'} \iint_A G(\xi, \eta)\,\bar{\chi}_R\bar{\chi}_\Psi\,d\xi d\eta + \frac{2\mu^4 m^2}{m^2 + \mu^2} U'^2 \iint_A G(\xi, \eta)\,\bar{\chi}_R\,d\xi d\eta\Big\} \qquad (16\,a)$$

Finally, substitution of the approximate autocorrelation functions described in Appendix I, and expansion of $E^\star$ in a Taylor series around $t = 0$ leads to Eq. (2) of Section IV.

Appendix III—The Predictability of Zonally Averaged Barotropic Flow

We now consider the MS error E in predictions of the zonal component $\bar{u}$ of the zonally averaged wind, starting with reconstructed initial wind fields in which R_u and R_v are the analysis errors in the eastward and northward wind components u and v.

$$E = \frac{1}{W}\int_0^W R^2 dy$$

where R is the error in $\bar{u}$ at various latitudes, and W is the entire width of the flow. Dif-

ferentiating E successively with respect to time,

$$\frac{\partial E}{\partial t} = \frac{2}{W}\int_0^W R\frac{\partial R}{\partial t}\,dy$$

$$\frac{\partial^2 E}{\partial t^2} = \frac{2}{W}\int_0^W \left[R\frac{\partial^2 R}{\partial t^2} + \left(\frac{\partial R}{\partial t}\right)^2\right] dy$$

$$\frac{\partial^3 E}{\partial t^3} = \frac{2}{W}\int_0^W \left[R\frac{\partial^3 R}{dt^3} + 3\frac{\partial R}{\partial t}\frac{\partial^2 R}{\partial t^2}\right] dy$$

$$\frac{\partial^4 E}{\partial t^4} = \frac{2}{W}\int_0^W \left[R\frac{\partial^4 R}{\partial t^4} + 4\frac{\partial R}{\partial t}\frac{\partial^3 R}{\partial t^3} + 3\left(\frac{\partial^2 R}{\partial t^2}\right)^2\right] dy \qquad (17)$$

Initially, R vanishes, because the analysis error R_u is random and oscillates around zero. Moreover, the first derivative of $\bar{u}$ is given by

$$\frac{\partial \bar{u}}{\partial t} = -\frac{\partial}{\partial y}\overline{uv}$$

in which a "bar" above a quantity now (and henceforth) denotes its *zonal* average. The equation above implies that, *initially*

$$\frac{\partial R}{\partial t} = -\frac{\partial}{\partial y}\left(\overline{vR_u} + \overline{uR_v} + \overline{R_uR_v}\right)$$

Since the analysis errors R_u and R_v are assumed to be uncorrelated with the true initial wind field, and are themselves independent, $\partial R/\partial t$ vanishes *initially* at all latitudes. Together with Eq. (17), these results imply that

$$\left(\frac{\partial E}{\partial t}\right)_0 = \left(\frac{\partial^2 E}{\partial t^2}\right)_0 = \left(\frac{\partial^3 E}{\partial t^3}\right)_0 = 0$$

$$\left(\frac{\partial^4 E}{\partial t^4}\right)_0 = \frac{6}{W}\int_0^W \left(\frac{\partial^2 R}{\partial t^2}\right)_0^2 dy$$

Thus, expanding E in a Taylor series around $t = 0$,

$$E \simeq \frac{t^4}{4W}\int_0^W \left(\frac{\partial^2 R}{\partial t^2}\right)_0^2 dy \qquad (18)$$

From this formula, it follows immediately that the MS error in the zonally-averaged wind increases like the *fourth* power of the period.

It remains to estimate the second time-derivative of the error R for any particular latitude at $t = 0$. This can be done most simply by making use of Eq. (21) of THOMPSON (1957), in an approximate form justified later in that article. At any latitude,

$$\frac{\partial^2 \bar{u}}{\partial t^2} \simeq \bar{\nu}^2\frac{\partial^2 U}{\partial y^2} - 2\overline{\nu\nabla^2\nu}\,U \qquad (19)$$

in which

$$U = \bar{u} - \frac{1}{W}\int_0^W \bar{u}\,dy$$

Now, since R vanishes *initially*, Eq. (19) implies that

$$\left(\frac{\partial^2 R}{\partial t^2}\right)_0 = \frac{\partial^2 U}{\partial y^2}\left[\overline{(\nu + R_\nu)^2} - \bar{\nu}^2\right]$$

$$- 2U\left[\overline{(\nu + R_\nu)\left(\nabla^2\nu + \nabla^2 R_\nu\right)} - \overline{\nu\nabla^2\nu}\right]$$

Thus, because R_ν is random with respect to ν,

$$\left(\frac{\partial^2 R}{\partial t^2}\right)_0 = \overline{R_\nu^2}\frac{\partial^2 U}{\partial y^2} - 2\overline{R_\nu\nabla^2 R_\nu}\,U$$

For purposes of estimate, we now introduce a characteristic wave-number m of the initial error field, and a characteristic wave-number $\overline{M}$ of meridional fluctuations in the mean zonal wind profile. At the same time recalling that the statistics of the initial error field are isotropic, we may rewrite the equation above in the form

$$\left(\frac{\partial^2 R}{\partial t^2}\right)_0 = \frac{1}{2}\left(2m^2 - \overline{M}^2\right)\overline{\mathbf{R}\cdot\mathbf{R}}\,U$$

where $\mathbf{R}$ is the vector error in the initial wind field. This equation, taken together with Eq. (18), leads to Eq. (3) of Section VI.

REFERENCES

BRIER, G. W., 1944: Unpublished manuscript. U. S. Weather Bureau, Washington, D. C.

CHARNEY, J. G., 1948: "On the scale of atmospheric motions". *Geofys. Publ.* **17,** No. 2.

EADY, E. T., 1952: "Note on Weather Computing and the So-Called 2 1/2 Dimensional Model". *Tellus* **4,** pp. 157—168.

FJØRTOFT, R., 1950: "Application of Integral Theorems in Deriving Criteria of Stability for Laminar Flows and for the Baroclinic Circular Vortex". *Geofys. Publ.* **17,** No. 6.

KUO, H.-L., 1956: "On Quasi-Nondivergent Prognostic Equations and Their Integration". *Tellus* **8,** pp. 373—383.

LERAY, J., 1931: "Sur le système d'équations aux derivatives partielles qui régit l'écoulement des fluides visqueux. *Comptes rendus de l'Academie des Sciences*, **192,** pp. 1180—1185.

LORENZ, E. N., 1953: "The Interaction between a Mean Flow and Random Disturbances". *Tellus* **5,** pp. 238—250.

NAMIAS, J., 1947: "*Extended Forecasting by Mean Circulation Methods*". U. S. Weather Bureau, Washington, D. C.

PHILLIPS, N. A., 1951: "A simple three-dimensional model for the study of large-scale extratropical flow patterns". *Jour. Met.* **8,** pp. 381—394.

PHILLIPS, N. A., 1956: "The general circulation of the atmosphere: a numerical experiment". *Quart. Jour. R.M.S.* **82,** pp. 123—164.

THOMPSON, P. D., 1953: "On the theory of large-scale disturbances in a two-dimensional baroclinic equivalent of the atmosphere". *Quart. Jour. R.M.S.* **79,** pp. 51—69.

THOMPSON, P. D., 1957: "A Heuristic Theory of Large-Scale Turbulence and Long-Period Velocity Variations in Barotropic Flow". *Tellus* **9,** pp. 69—91.

Experiments in Numerical Forecasting, Using Space-Smoothed Fields*

By HLYNUR SIGTRYGGSSON[1] and AKSEL WIIN-NIELSEN[2],
International Meteorological Institute in Stockholm

(Manuscript received February 28, 1957)

Abstract

A number of experiments in numerical forecasting using a technique based on the theory developed by Fjörtoft have been performed.

Section A contains the results of making barotropic 500 mb forecasts for 24 and 48 hours where the vorticity is advected in a space-smoothed field held constant in 12 or 24 hours. The results indicate that the forecasts in the mean are slightly less accurate as measured by the correlation coefficient between computed and observed changes. It is pointed out that the root mean square of the computed changes is smaller than for ordinary barotropic forecasts, especially for a large area.

Section B contains the results of an experiment to forecast the thickness field between 1,000 and 500 mb. The forecasts are based on the theory developed by FJÖRTOFT (1955). The flow at 500 mb has, however, in these forecasts been forecast with the barotropic model. The errors in these forecasts are of the same order of magnitude as in the barotropic forecasts. In the last part it is tried to relate the errors to physical defects of the model, especially the assumption of adiabatic motion.

Introduction

The graphical methods, developed by FJÖRTOFT (1952 and 1955), by which barotropic prognoses can be made, rest upon a number of assumptions necessary in order to make graphical technique available to the forecasting problem. The basic equation is the barotropic vorticity equation, but in addition to the assumptions about the behaviour of the atmosphere, which have to be made in order to arrive at this simple forecasting equation, two main approximations are necessary, one of a physical and the other one of a mathematical nature.

The physical assumption is that a field obtained by a certain space smoothing of the contour field can be assumed to remain constant during a time T which is large compared with the time step generally used in numerical forecasting. The nature of the space smoothing is such, that a certain wavelength is completely eliminated, while all other wavelengths are damped. The greater the wavelength is the smaller will be the damping. In general the smoothing operator must be

* The research reported in this document has been sponsored partly by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, United States Air Force, under contract No. AF 61(514)—648-C, through the European Office ARDC. and partly by The Swedish Board of National Defense under contract FOA 2329-299, 1956.

[1] while on leave from the Icelandic Meteorological Service.

[2] on leave from the Danish Meteorological Institute.

chosen in such a way that the wavelengths, which cause the most rapid change in the unsmoothed field, are eliminated. How large T may be taken can only de determined from experience. Fjörtoft has shown that a larger T can be taken, if the simple advection in the space-smoothed field is replaced by two successive displacements, the first one performed essentially in the vorticity field of the space-smoothed height field, the other one in a double-smoothed field. The forecasts presented in this investigation have, however, been computed using a displacement in a field smoothed only once. If this simple displacement rule is used, one can hardly take T greater than 24 hours. The main part of the forecasts here are made using $T=12$ hours. A few of the forecasts have been recomputed using $T=24$ hours. We shall later comment on the differences between these two sets of forecasts.

The mathematical approximation appears in the solution of the Poisson equation, which must be solved, when the vorticity has been displaced in the space-smoothed field. The approximative solution used (FJÖRTOFT 1952) may cause rather great errors, especially for the larger wavelengths.

It is usually very difficult to distinguish between errors introduced by the first and the second approximation. Because of this it is of some interest to make a number of forecasts in which only one of the approximations is made and compare the result with forecasts made without any of the two approximations mentioned. The goodness of the approximative solution of the Poisson equation is most easily tested by computing the finite difference form of the "Laplacian" of a certain field and use the approximative solution to recompute the original field. A test of this kind has already been performed by FJÖRTOFT (1952). The goodness of keeping the smoothed field constant during the time T has, however, not been tested separately. We have therefore arranged the following computations in such a way, that the Poisson equation is solved "exactly", while the advection is performed in the smoothed field held constant for a certain time. The "exact" solution of the Poisson equation can in practice only be obtained with an electronic computer. The whole forecast was therefore arranged for a computer. For computational reasons it was convenient to make minor changes in the procedure used when applying the graphical technique. The results obtained in this paper are not directly comparable with results obtained by a graphical technique. Recent investigations in Norway and Germany have shown that the finite difference methods used in a grid representation may introduce a false dispersion, because the longer harmonic components would move faster than the shorter ones. Our computations have not been corrected for this effect.

The second part of this paper contains the result of a number of forecasts made in order to test a method, by which a prognosis of the thickness field between two isobaric surfaces can be made (FJÖRTOFT 1955). We have here chosen the thickness field between 1,000 and 500 mb. This method has also been developed by Fjörtoft in such a way that it is suitable for graphical methods. It was shown that a certain quantity derived from the thickness field is individually conserved in a smoothed field. The latter field is again assumed time independent for a period of $T=$ 24 hours. In the computations reported here it was decided *not* to make the last assumption. Instead it was decided to incorporate the thickness forecast in the ordinary barotropic forecast. Thus the thermal quantity was advected geostrophically in the height field of the 500 mb surface, forecast from the *barotropic* model.

Besides the interest in the accuracy of any numerical forecasts, one must also consider the practicability of such forecasts, for example, the computation time. The barotropic forecasts as well as the thickness forecasts presented here require a smaller computation time than is generally used in forecasting with one or two parameter models. The gain in computation time is in both methods obtained by solving the differential equations only a few times compared with the numerical techniques generally used. The results show that the prognostic value of the barotropic forecasts is almost the same. It has not been possible to compare the thickness forecasts with forecasts obtained from another model. The thickness forecasts have, however, a definite prognostic value as is shown in the examples in section B.

A. Barotropic prognoses

1. The computational procedure

The basic equation for geostrophic, barotropic forecasts can be written:

$$\frac{\partial \xi}{\partial t} = \frac{1}{f} J(\eta, \phi) \tag{1}$$

where ϕ is the geopotential, ξ the relative vorticity computed from the geostrophic assumption as $\xi = \frac{1}{f} \nabla^2 \phi$ and $\eta = \xi + f$; f denotes the Coriolis parameter. In finite difference form eq. (1) becomes:

$$\frac{\partial \overline{\nabla}^2 \phi}{\partial t} = \frac{1}{4} J(\eta, \phi) = \frac{1}{4} J\left(\frac{m^2}{f} \frac{1}{(\Delta s)^2} \overline{\nabla}^2 \phi + f, \phi\right) \tag{2}$$

where $\overline{\nabla}^2$ and J denote the usual finite difference approximations of the Laplacian and the Jacobian, m the map magnification factor and Δs the grid size.

The smoothing operator is defined as follows:

$$\phi^{(N)} = \phi + k \overline{\nabla}^2 \phi \tag{3}$$

where $\phi^{(N)}$ denotes the smoothed value, while k is a constant. The effect of the smoothing operator on waves of a certain fixed wavelength is determined partly by the numerical value of k and partly by the grid size used to evaluate the finite difference form of the Laplacian. The smoothing operator (3) corresponds to the one used by Fjörtoft for $k = \frac{1}{4}$. This value is obviously the most convenient for graphical methods. Furthermore $k = \frac{1}{4}$ is a maximum value in the sense, that if k is taken larger than $\frac{1}{4}$ some of the short waves will be amplified. This can be seen, if we assume that the geopotential field can be described by the Fourier series:

$$\phi = \sum \phi_{mn} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} \tag{4}$$

The smoothed value of ϕ is then given by

$$\phi^{(N)} = \sum \left\{ (1 - 4k) + 2k\left(\cos \frac{\Delta s}{\lambda_m} 2\pi + \cos \frac{\Delta s}{\omega_n} 2\pi\right) \right\} \phi_{mn} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} \tag{5}$$

The amplitude of a certain wave is therefore changed by a factor A, which for equal values of λ_m and ω_n, i.e. $\lambda_m = \omega_n = \lambda$ takes the form

$$A = (1 - 4k) + 4k \cos \frac{\Delta s}{\lambda} 2\pi \tag{6}$$

It is easily seen from (6) that $A < -1$, if $k > \frac{1}{4}$ and $\lambda = 2\Delta s$. Values of $k > \frac{1}{4}$ should therefore not be used if the smoothing is performed by a numerical procedure. The possibility still exists to use such values in graphical work because the manual drawing of the isolines will automatically remove the undesired small waves from the field. The value actually used for k in (3) will in each case depend on which wavelengths we want to retain in the actual field. We shall return to this question later.

The smoothed field $\phi^{(N)}$ can be incorporated in the prognostic equation (2) by inserting from (3). We obtain then:

$$\frac{\partial \overline{\nabla}^2 \phi}{\partial t} = \frac{1}{4} J(\eta, \phi^{(N)}) - \frac{1}{4} J\left(\frac{m^2}{f} \frac{1}{(\Delta s)^2} \overline{\nabla}^2 \phi,\ k \overline{\nabla}^2 \phi\right) - \frac{1}{4} J(f,\ k \overline{\nabla}^2 \phi) \tag{7}$$

If we in accordance with Fjörtoft (1952) suppose that the relative variation in $\frac{m^2}{f} \frac{1}{(\triangle s^2)}$ is small compared with the relative variation in $\overline{\nabla}^2 \phi$ the second term on the right side disappears. With regard to the last term in (7) it has already been noted by Fjörtoft that the main contribution from the variability of the Coriolis parameter comes from the advection in the smoothed field. This effect is included in the first term, while the effect expressed by the last term is of minor importance. We can also compare the order of magnitude of the neglected second term with the contribution from the last term. The neglected term is:

$$\frac{1}{4} k \overline{\nabla}^2 \phi \frac{\partial \overline{\nabla}^2 \phi}{\partial x} \frac{\partial}{\partial y} \left(\frac{m^2}{f} \frac{1}{(\Delta s)^2}\right)$$

The magnitude of this term should be compared with the term

$$\frac{1}{4} k \frac{\partial \nabla^2 \phi}{\partial x} \frac{\partial f}{\partial y}$$

By taking the ratio between the two terms it is easily seen that the former one gives a greater contribution than the latter in regions, where the numerical value of $\nabla^2\phi$ exceeds a rather low limit. If it is allowed to neglect the second term in (7), we are therefore also allowed to disregard the third one. When this is done, the prognostic equation assumes the simple form:

$$\frac{\partial \nabla^2\phi}{\partial t} = \frac{1}{4} J(\eta, \phi^{(N)}) \qquad (8)$$

In using eq. (8) it must finally be decided which value should be used for the grid size and the factor k. The most common grid size used so far has been of the order of magnitude 300 km, while the value applied in the graphical methods has been around 600 km or even greater. The smaller grid size is in general desirable because the finite difference expressions then give a better approximation to the differentials. The choice of the grid size is, however, in these computations also influenced by the fact that we want a smoothed field $\phi^{(N)}$, which with a good approxination may be assumed constant for a rather long time. The smoothing operator should therefore be used in such a way that it removes those waves, which have the greatest influence on the time variation of the height field, and which therefore influence to the greatest extent the trajectory of the particles. Fjörtoft has adopted the value $L \sim 2{,}500$ km as the wavelength of the most rapid changing component of the height field. An estimate of the wavelength, which should be removed, can be made, if we assume that the undesired waves are those which in the time T move one fourth of the wavelength or more. From the well known Rossby formula

$$c = U - \frac{\beta L^2}{4\pi^2} \qquad (9)$$

we get by multiplication by T the displacement in the time interval T. Putting $cT = \frac{L}{4}$ we get the following equation for the determination of L:

$$\frac{\beta T}{4\pi^2} L^2 + \frac{1}{4} L - UT = 0 \qquad (10)$$

Table I contains values of L satisfying (10) for given characteristic values of U and using a value of $\beta = \frac{df}{dy}$ corresponding to 45° latitude.

Table I

U m sec^{-1}	10	20	30
L km for $T = 12$ h	1,560	2,880	4,020
L km for $T = 24$ h	2,550	4,310	5,750

The value used by Fjörtoft corresponds to a rather low velocity, when $T = 24$ hours, while $L \sim 2{,}500$ km will be rather good for $T = 12$ hours, because $U \sim 20$ m/sec^{-1} certainly is a more representative value than $U \sim 10$ m/sec^{-1} for the 500 mb surface. From the result of the estimate it was therefore decided to use $T = 12$ hours and try to arrange the smoothing in such a way that wavelengths around 2,400 km were most effectively smoothed. With a grid size of 300 km, which is desirable for an accurate determination of the vorticity, we can, however, not obtain such a smoothing, because the maximum wavelength, which can be removed, is 1,200 km corresponding to $k = \frac{1}{4}$. It was therefore decided to use $\Delta s \sim 600$ km and $k = \frac{1}{4}$ when computing the smoothed field, but retain $\Delta s = 300$ km for the computation of all other quantities. The slight inconsistency introduced hereby is not considered serious because of the approximate proportionality between the two finite difference forms of the Laplacian.

The forecast was then in principle made by advecting the absolute vorticity in the smoothed field kept constant for 12 hours using a time step $\Delta t \sim 1$—1 ½ hours in the advection. After having completed the advection, the Poisson equation was solved to give the height field ϕ^{12}. This procedure was then repeated to give a 48-hour forecast. Two of the forecasts have been repeated using $T = 24$ hours, but with no change in the smoothing operation. The Poisson equation was solved by the extrapolated Liebmann method, using the known height field from 12 hours earlier as a first guess. The solution of the Poisson equation needs a great number of iterations, normally 50—60, because the first guess is

rather poor. Attempts were made to decrease the number of iterations by solving the Poisson equation already after the first time step and to use the value ϕ^1 obtained in this way to make a linear extrapolation and to compute an approximative value ϕ_a^{12}, which could be used as the starting value in the iterative procedure. The number of iterations decreased only slightly by this procedure, presumably because the forecast height change over an interval of this length is not linear. Much more time was saved by increasing the over-relaxation factor to the full Fraenkel-coefficient instead of the arithmetic mean between the Fraenkel-coefficient and the Richardson-coefficient generally used.[1] By replacing the overrelaxation factor $\alpha = 0.33$ generally used by $\alpha_F = 0.45$, the convergence was more than twice as fast.

Before discussing the result of the computations it should finally be noted that the forecasts were made over a region corresponding to the one used in the barotropic prognoses reported by Döös (1956), and the same statistical quantities as used by him have been computed to get a measure of the goodness of the forecasts. The time required to compute a forecast for 24 hours using $T = 12$ hours, and an optimum over-relaxation is 14—16 min. against approximately 25 min. for the standard procedure computing the height field in each time step. These forecasts will hereafter be denoted B-forecasts, while those obtained by using a smoothed advection field will be denoted S-forecasts.

3. Results of the barotropic computations

The forecasts have been made from numerically analysed maps in order to facilitate the comparison between these forecasts and the B-forecasts computed on a routine basis by the Royal Swedish Air Force.

A visual comparison between the two sets of forecasts shows first of all that the patterns in general agree very well with each other. One has to go into details in order to find differences between the prognoses. The analysis from 7. 2. 1955 15 GMT and the 24 and 48 hours forecasts using eq. (8) together with the corresponding verification maps

[1] The authors are indebted to Mr. A. Bring for pointing out this possibility of speeding up the relaxation procedure.

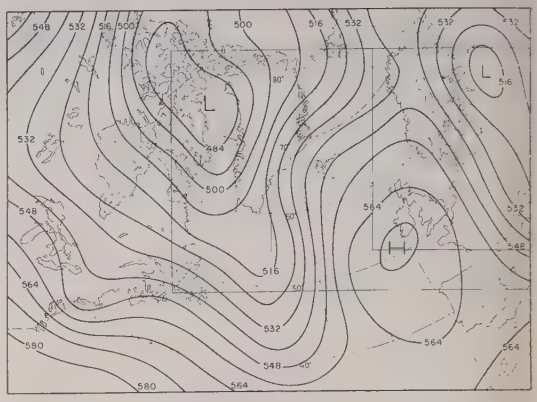


Fig. 1 a. 500 mb numerical analysis from 7. 2. 1955 15 GMT. The lines are here as in the figures 1 to 6 drawn with an interval of 8 dam. The thin solid lines denote the regions of verification.

are reproduced in fig. 1 a—e. The dashed lines on the two forecasts are isolines for the difference in height between the forecast from eq. (8) and the conventional barotropic forecast. One observes that this difference has a quite large scale, the forecasting area being divided essentially in a positive and a negative region. It is not easy to explain the distribution of these differences in detail. They have a maximum of 180 m in 24 hours and 370 m in 48 hours. There exist the two possibilities that they may be caused by the advection in the space mean field during the rather long time interval ($T = 12$ h) or may be ascribed to the different numerical treatment used in the two cases. Although the error introduced by the assumption of the time independence of the space mean field may be rather large, it is not probable that it in general will result in differences between the two kinds of prognoses which is of the same sign over large regions. On the contrary one would expect regions of positive and negative differences strongly related to the actual motion of the waves.

Several of the small scale features in the field representing the difference between the S- and B-forecasts may be ascribed to actual movement of the waves, a motion, which has not been considered because of the assumed constancy in time of the smoothed field, but the large scale nature of the difference field must be explained in another way.

The large scale pattern in the difference field suggests that the circulation around the

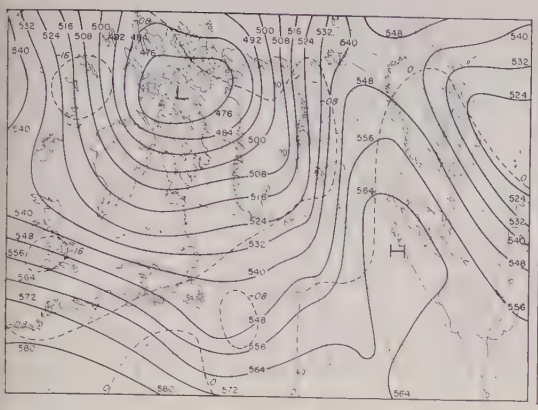


Fig. 1 b. 24 hour barotropic S-forecast, based on the analysis shown in fig. 1 a. The dashed lines show the difference between the S and the B forecasts.

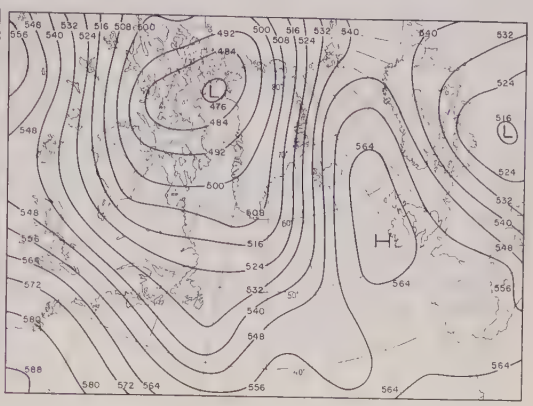


Fig. 1 d. Numerical analysis from 8. 2. 1955 15 GMT.

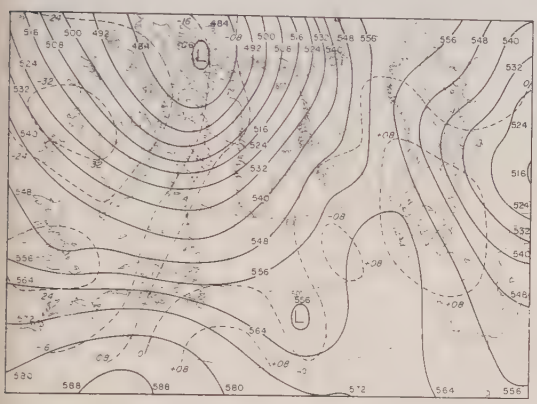


Fig. 1 c. 48 hour barotropic S-forecast, based on the analysis shown in fig. 1 a.

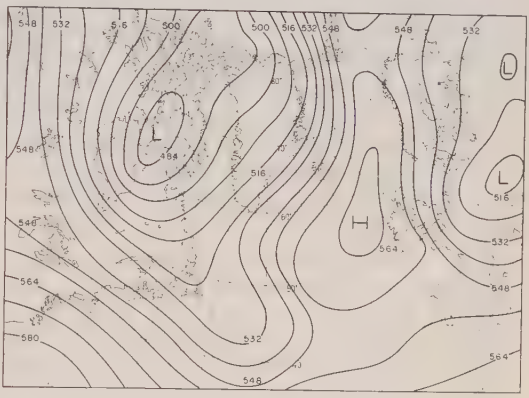


Fig. 1 e. Numerical analysis from 9. 2. 1955 15 GMT.

forecasting region is different in the two methods. We can estimate the difference in the circulation for the two methods in the following way, where we for simplicity have assumed a strictly non-divergent flow. Denoting the conventional barotropic forecast by the subscript B and the other by S, we get:

$$\frac{\partial}{\partial t}\Big\{ \oint (\mathbf{v}_B - \mathbf{v}_S) \cdot \delta \mathbf{s} \Big\} =$$

$$= \frac{\partial}{\partial t} \Big\{ \int_A \big(\nabla^2 \psi_B - \nabla^2 \psi_S \big)\, dA = \Big.$$

$$= \int_A \big\{ J\big(\nabla^2 \psi + f,\ \psi - \psi^{(N)} \big) \big\}\, dA$$

The rate of change of the circulation has here been transformed using Stokes' theorem and the two forecasting equations. With the aid of the definition of the smoothing operator (FJÖRTOFT 1955) we can write:

$$\psi - \psi^{(N)} = -\frac{1}{a_N} \nabla^2 \psi$$

When this definition is introduced we are left with the following expression for the rate of change of the circulation:

$$-\frac{1}{a_N} \int_A J(f,\ \nabla^2 \psi)\, dA$$

This surface intregral is then transformed to an integral along the boundary. We obtain

$$-\frac{1}{a_N} \int_A J(f,\ \nabla^2 \psi)\, dA = \frac{1}{a_N} \oint \nabla^2 \psi \frac{\partial f}{\partial s}\, ds$$

and consequently:

$$\frac{\partial}{\partial t}\left\{\oint (\mathbf{v}_B - \mathbf{v}_s)\cdot \delta \mathbf{s}\right\} = \frac{1}{a_N}\oint \nabla^2\psi \frac{\partial f}{\partial s} ds$$

The last equation states that the difference in the rate of change of circulation in the two forecasting systems depends upon the correlation between the relative vorticity and the change of the Coriolis parameter along the boundary. If there is a positive correlation between $\nabla^2\psi$ and $\frac{\partial f}{\partial s}$, we get

$$\frac{\partial}{\partial t}\oint \mathbf{v}_B \cdot \delta \mathbf{s} > \frac{\partial}{\partial t}\oint \mathbf{v}_s \cdot \delta \mathbf{s}$$

while the opposite holds if $\nabla^2\varphi$ and $\frac{\partial f}{\partial s}$ are negatively correlated along the boundary. The equation above may express one of the mechanisms responsible for the large scale pattern in the difference between the two forecasts. The possibility that the large differences may be ascribed to the different numerical treatment was tested in the following way. The main difference in the computations is that a Poisson equation is solved in every time step in the B-forecast, while the corresponding equation is solved only once every 12 hours in the S-forecasts. Each time the equation is solved in the approximative way a certain error is introduced. The possibility exists that these errors may accumulate during the forecast in certain regions of the map and give rise to the great differences observed between the two forecasts. An estimate of the error made in solving the difference equation by a Liebmann procedure is, however, difficult and laborious. The B-forecast for 7. 2. 1956 15Z was therefore recomputed in such a way that the Poisson equation was solved more accurately by reducing the tolerance and increasing the over-relaxation factor to its maximum value. No significant change was, however, observed. It should be noted that a distribution of differences similar to the case described above was observed in several of the other cases computed. The general rise of the 500 mb surface often observed in the B-forecasts is not so predominant in the S-forecasts.

The results of the forecasts are given in table II and III. The correlation coefficient between the observed and computed changes is denoted by r; σ_x and σ_y are the root mean squares of the observed and computed changes and ε is the root mean square of the difference between the observed and computed changes. The verification was made over a small and a large region as indicated by the thin solid squares on fig. 1 a. Table II gives the values for the small and table III for the large region.

Tables II and III show that the goodness of the S forecasts is comparable with the goodness of the B-forecasts as measured by the correlation coefficient. As expected, the mean values of r is a little lower for the S-forecasts, although several cases exist, in which it is greater. The rather small difference between the two correlation coefficients is but another expression for the fact that the forecasts over the verification region are quite similar. The extreme values of the differences of the kind mentioned above happened to be outside the large verification region in most cases.

The root mean square of the predicted changes is in the mean smaller for the S-forecasts than in the other case. This is true for 24 and 48 hours forecasts in the large as well as the small region. In the individual cases it happens only a few times that the opposite is true. It has earlier been noted by Döös (1956) that the predicted changes are in general greater than the observed changes ($\sigma_y > \sigma_x$) for the B-forecasts. The S-forecasts represent therefore an improvement in this respect as far as one can judge from the few cases computed.

The mean error (ε) for the predicted changes has as the other quantities the same order of magnitude for the two sets of forecasts. From the mean values of ε and r it is seen that the assumed time independence of the smoothed field has no significant influence on the goodness of the forecasts. The larger mean error in certain of the individual cases may, however, be ascribed to larger changes in the smoothed field. This is especially true for the case from 20. 2. 1955 where a marked change of the smoothed field occurred over the Atlantic.

Tables IV and V contain the result of the recomputation of two cases, where T was put equal to 24 hours instead of 12 hours. The

Table II. $T = 12$ hours, small region

Date		24-hour forecast					48-hour forecast				
		r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x
30/1-56 15	B	0.96	85	94	36	0.43	0.87	127	194	150	1.18
	S	0.93	85	89	31	0.36	0.86	127	128	78	0.61
31/1-56 15	B	0.72	64	82	73	1.14	0.44	109	147	189	1.73
	S	0.70	64	68	58	0.90	0.58	109	97	143	1.31
1/2-56 15	B	0.77	74	57	63	0.86	0.57	119	105	164	1.38
	S	0.57	74	56	75	1.01	0.62	119	93	145	1.22
2/2-56 15	B	0.78	76	67	47	0.62	—	—	—	—	—
	S	0.81	76	77	46	0.61	—	—	—	—	—
7/2-56 15	B	0.87	74	75	38	0.52	0.87	125	160	79	0.64
	S	0.88	74	61	35	0.47	0.85	125	121	78	0.62
8/2-56 15	B	0.88	96	115	64	0.66	0.77	144	198	131	0.91
	S	0.90	96	113	75	0.78	0.70	144	189	180	1.25
9/2-56 15	B	0.59	84	76	69	0.82	—	—	—	—	—
	S	0.24	84	64	99	1.18	—	—	—	—	—
20/2-56 15	B	0.88	82	86	58	0.70	0.74	84	101	86	1.02
	S	0.79	82	106	96	1.17	0.57	84	195	204	2.43
Mean	B	0.81	79	82	56	0.72	0.71	118	151	133	1.14
	S	0.73	79	79	64	0.81	0.69	118	137	136	1.24

Table III. $T = 12$ hours, large region

Date		24-hour forecast					48-hour forecast				
		r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x
30/1-56 15	B	0.85	90	119	63	0.72	0.66	119	191	146	1.23
	S	0.82	90	107	65	0.76	0.52	119	181	159	1.34
31/1-56 15	B	0.42	75	101	98	1.31	0.18	111	147	167	1.50
	S	0.29	75	91	98	1.31	0.11	111	145	172	1.55
1/2-56 15	B	0.32	72	97	101	1.40	0.14	102	123	145	1.42
	S	0.32	72	86	93	1.29	0.20	102	126	140	1.37
2/2-56 15	B	0.73	89	108	75	0.84	—	—	—	—	—
	S	0.80	89	91	56	0.63	—	—	—	—	—
7/2-56 15	B	0.67	84	125	97	1.16	0.68	121	208	162	1.34
	S	0.75	84	95	64	0.77	0.69	121	162	131	1.08
8/2-56 15	B	0.85	94	144	97	1.03	0.81	162	261	174	1.07
	S	0.83	94	115	79	0.84	0.77	162	249	195	1.20
9/2-56 15	B	0.82	103	150	94	0.91	—	—	—	—	—
	S	0.63	103	113	96	0.95	—	—	—	—	—
20/2-56 15	B	0.65	74	105	91	1.23	0.64	99	172	157	1.59
	S	0.49	74	103	105	1.42	0.48	99	227	237	2.39
Mean	B	0.66	85	119	90	1.08	0.52	119	184	159	1.36
	S	0.62	85	100	82	1.00	0.46	119	182	172	1.39

Table IV. $T = 24$ hours, small region

Date		24-hour forecast					48-hour forecast				
		r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x
7/2-56 15	B	0.87	74	75	38	0.52	0.87	125	160	79	0.64
	S	0.84	74	57	40	0.54	0.80	125	133	117	0.94
8/2-56 15	B	0.88	96	115	64	0.66	0.77	144	198	131	0.91
	S	0.83	96	119	98	0.98	—	—	—	—	—

Table V. $T = 24$ hours, large region

Date		24-hour forecast					48-hour forecast				
		r	σ_x	σ_y	ε	r/σ_x	r	σ_x	σ_y	ε	ε/σ_x
7/2-56 15	B	0.64	84	125	97	1.16	0.68	121	208	162	1.34
	S	0.67	84	99	75	0.90	0.69	118	155	131	1.11
8/2-56 15	B	0.85	94	144	97	1.03	0.81	162	261	174	1.07
	S	0.71	94	111	88	0.93	—	—	—	—	—

selected cases are two of the better ones. As expected, there is if anything, an increase in the mean error and a decrease in the correlation coefficient. The change in ε and r is, however, only very small in the two selected cases, and the prognostic value of the two forecasts with $T = 24$ hours is quite as good as when $T = 12$ hours. No conclusions can naturally be drawn from these two cases, but if further investigations should be done with a computational scheme of this kind experiments should be made to find the most suitable value for T.

4. Conclusions

The results of the computations show that the S-forecasts are in general comparable with the B-forecasts. This is true, not only as summarized in the tables, but is also the general impression from a visual comparison of the two sets of forecasts. The value of T has been taken rather small as compared to the value generally used when graphical technique is applied. A few tests have, however, shown that a larger value of T may possibly be used without too much loss in accuracy. Further investigations are needed along this line before final conclusions can be drawn.

B. Thickness Forecast

1. Review of Theory

A forecast of a thickness field might be useful in several ways even when made with the usual barotropic model. It is a simple baroclinic model, and could possibly, because of its simplicity, be a useful tool in analysing the errors encountered with a more advanced baroclinic forecast. It could, together with a barotropic forecast for the 500 mb level, serve as means for obtaining, by interpolation, forecasts for other levels. Furthermore, it could be used as a preliminary field for making an objective thickness analysis by the method of Bergthorsson and Döös.

An experiment was therefore made to use the Swedish electronic computer (BESK) to obtain a forecast for the thickness between two isobaric surfaces, in accordance with the theory given below.

FJÖRTOFT (1955) has described a method, whereby a numerical forecast for the thickness of an atmospheric layer can be made. We shall in the following briefly describe the derivation of the conservation theorem used for the thickness forecast and mention the

different assumptions made in order to obtain this simple equation. If not otherwise stated, the notations are the same as used by FJÖRTOFT.

1. The vorticity equation is used in the form:

$$\frac{\partial \eta}{\partial t} = -\mathbf{v} \cdot \nabla \eta + f \frac{\partial w}{\partial z} \qquad (11)$$

where the vertical advection of vorticity, the "twisting term" and the solenoidal term have been neglected. In the divergence term η has been replaced by f and $-\operatorname{div}_2 \mathbf{v}$ by $\frac{\partial w}{\partial z}$, using the continuity equation in the form:

$$\operatorname{div}_3 \mathbf{v} = 0$$

2. Adiabatic conditions are assumed in the thermodynamic equation, i.e.

$$\frac{d \ln \Theta}{dt} = \frac{\partial \ln \Theta}{\partial t} + \mathbf{v} \cdot \nabla \ln \Theta + ws = 0 \qquad (12)$$

where $s = \frac{\partial \ln \Theta}{\partial z}$ is a measure of the vertical stability, which is assumed constant in time and space.

3. The thermal wind equation is used in the simplified form:

$$\frac{\partial \mathbf{v}}{\partial z} = -\frac{g}{f} \nabla \ln \Theta \times \mathbf{k} \qquad (13)$$

4. The horizontal wind $\mathbf{v}$ is assumed non-divergent and described by a stream-function ψ, i.e.

$$\mathbf{v} = -\nabla \psi \times \mathbf{k} \qquad (14)$$

ψ is further assumed to vary linearly with height, i.e.

$$\frac{\partial^2 \psi}{\partial z^2} = 0 \qquad (15)$$

which implies that the stream-function can be written as

$$\psi = \hat{\psi} + \frac{z}{h} \psi_T \qquad (16)$$

$2h$ is the "effective" depth of the atmosphere. z is the vertical coordinate counted from the middle of the atmosphere, $\hat{\psi}$ is the stream function at $z = 0$, and

$$\psi_T = \frac{1}{2} (\psi_{z=h} - \psi_{z=-h}) \qquad (17)$$

It follows further that the horizontal wind can be written

$$\mathbf{v} = \hat{\mathbf{v}} + \frac{z}{h} \mathbf{v}_T \qquad (18)$$

Inserting from (13) and (14) in (12) we obtain:

$$\frac{\partial}{\partial t} \left(\frac{\partial \psi}{\partial z} \right) = -\hat{\mathbf{v}} \cdot \nabla \frac{\partial \psi}{\partial z} - \frac{g}{f} s w \qquad (19)$$

This equation is integrated with respect to z from $-h$ to $+h$, and we obtain after dividing by 2:

$$\frac{\partial \psi_T}{\partial t} = -\hat{\mathbf{v}} \cdot \nabla \psi_T - \frac{1}{2} \frac{g}{f} s \int_{-h}^{h} w dz \qquad (20)$$

The integral in the last term in eq. (20) is now evaluated in the following way: An equation for the vertical velocity is obtained by elimination of the time derivatives between the vorticity equation and the adiabatic equation. By differentiation of the vorticity equation with respect to z and taking the horizontal "Laplacian" of the adiabatic equation we obtain by subtraction of the two equations and neglecting horizontal variations of f:

$$f \frac{\partial^2 w}{\partial z^2} + \frac{g}{f} s \nabla^2 w =$$

$$= \frac{\partial}{\partial z} (\mathbf{v} \cdot \nabla \eta) - \nabla^2 \left(\hat{\mathbf{v}} \cdot \nabla \frac{\partial \psi}{\partial z} \right) \qquad (21)$$

This partial differential equation can be transformed by the introduction of Fjörtoft's smoothing operator expressing $\nabla^2 w$ in the following way:

$$\nabla^2 w = a_N (w^{(N)} - w) \qquad (22)$$

An expression for $w^{(N)}$ is obtained from the adiabatic equation:

$$\frac{g}{f} s w^{(N)} = -\frac{\partial}{\partial t} \left(\frac{\partial \psi}{\partial t} \right)^{(N)} - \left(\hat{\mathbf{v}} \cdot \nabla \frac{\partial \psi}{\partial z} \right)^{(N)} \qquad (23)$$

Inserting from (22) and (23) in (21) we finally get the following equation for w:

$$f\frac{\partial^2 w}{\partial z^2} - \frac{gs}{f} a_N w = F_1 + zF_2 \qquad (24)$$

where

$$F_1 = a_N\left[\frac{D}{Dt}\left(\frac{\partial \psi}{\partial z}\right)^{(N)} + (\hat{\mathbf{v}} - \hat{\mathbf{v}}^{(N)}) \cdot \nabla \frac{\partial \psi}{\partial z}\right]$$

and

$$F_2 = a_N\left(\frac{\partial \mathbf{v}}{\partial z} - \frac{\partial \mathbf{v}^{(N)}}{\partial z}\right) \cdot \nabla \frac{\partial \psi}{\partial z}$$

The equation for w can now be solved as an ordinary second order differential equation, because F_1 and F_2 according to the model assumptions are independent of height. The general solution is:

$$w = -\frac{1}{fr^2} F_1 - \frac{1}{fr^2} F_2 z + C_1 e^{rz} + C_2 e^{-rz} \qquad (25)$$

where $r^2 = \frac{gsa_N}{f^2}$. C_1 and C_2 are integration constants, which are determined from the boundary conditions $w = 0$ for $z = -h$ and $z = +h$.

With the expression (25) for w we can evaluate the integral $\int_{-h}^{h} w dz$ and insert the result in (20). We obtain

$$\frac{D\hat{}}{Dt}(\psi_T - k_N \psi_T^{(N)}) = k_N(\hat{\mathbf{v}} - \hat{\mathbf{v}}^{(N)}) \cdot \nabla \psi_T \qquad (26)$$

where

$$\frac{D\hat{}}{Dt} = \frac{\partial}{\partial t} + \hat{\mathbf{v}} \cdot \nabla$$

and

$$k_N = 1 - \frac{\operatorname{tgh} \lambda}{\lambda}; \quad \lambda^2 = r^2 h^2 = \frac{ga_N s}{f^2} h^2 \qquad (27)$$

If the term on the right side of (26) is neglected, which means that we neglect the advection of the thickness lines in the vorticity field at the mean level, we have an equation expression the conservation of the quantity $\xi = \psi_N - k_N \psi_T^{(N)}$ in the *vertically averaged flow.*

Another conservation theorem can, however, be obtained by transforming (26) to the following equation

$$\frac{D^N}{Dt}(\psi_T - k_N \psi_T^{(N)}) =$$

$$= -(\hat{\mathbf{v}} - \hat{\mathbf{v}}^{(N)}) \cdot \nabla [(1 - k_N)\psi_T - k_N \psi_T^{(N)}] \qquad (28)$$

The term on the right side of (28) expresses the advection of the quantity $[(1 - k_N)\psi_T - k_N \psi_T^{(N)}]$, which is closely related to the thermal vorticity, in the vorticity field at the mean level.

The neglection of the right hand side of (28) is probably a much better assumption than neglecting the corresponding term in (26), because the vorticity field as well as the quantity related to the thermal vorticity have a rather small scale, while the thickness field has a large scale compared with the scale of the vorticity field.

The following two conservation theorems can therefore be obtained from (26) and (28):

$$\frac{D\hat{}}{Dt}(\psi_T - k_N \psi_T^{(N)}) = 0 \qquad (29)$$

$$\frac{D^N}{Dt}(\psi_T - k_N \psi_T^{(N)}) = 0 \qquad (30)$$

of which (30) probably is the better one. The difference in validity between (29) and (30) can, however, only be judged by actually making forecasts using both equations on the same data. This has been done, and we shall return to the results later.

The value of k_N, given by (27), must be taken as an average value. In accordance with Fjörtoft we have chosen $k_N = \frac{1}{2}$, corresponding to $h = \frac{1}{2} \cdot 10^4$ m, $f = 10^{-4}$ sec^{-1}, $\frac{\partial \ln \Theta}{\partial z} = 10^{-5}$ m^{-1} and $a_N \sim \frac{4}{(\Delta s)^2} = 10^{-11}$ m^{-2}, where $\Delta s = 600$ km is the grid size. We shall later discuss the validity of assuming k_N to be a constant in space and time and the sensitivity of the thickness forecast to the value chosen for k_N.

Computational procedure

The surfaces chosen were the ones for 1,000 and 500 mb. Winds and vorticity were computed geostrophically. In the equations (29) and (30), ψ_T may, with a slight approximation, be replacedby ϕ_T, the geopotential thickness of the layer 1,000 to 500 mb. Some

of the operations described later required a grid with 600 km mesh size at 50° N, but unless otherwise stated, the 300 km mesh was used in the computations. After slight initial smoothing of the data fields with smoothing factor $\frac{1}{16}$, ϕ_T was calculated as the difference between the height fields of the 500 and 1,000 mb surfaces. The quantity ξ was then calculated according to the formula,

$$\xi = \phi_T - \frac{1}{2}\phi_T^{(N)} = \frac{1}{2}\phi_T - \frac{1}{8}\nabla^2\phi_T \qquad (31)$$

A 600 km grid size was used in order to satisfy the condition that $k_N = \frac{1}{2}$. Since ξ cannot be calculated at points in the two lines and columns closest to the boundary, the computational area had to be reduced from 30×39 grid points normally used in the barotropic forecast, to an area of 26×35 points.

The field in which ξ was advected was obtained from a barotropic forecast, made according to the method described by Staff Members, Institute of Meteorology (1954). The additional computations made in a general time step in order to forecast the thickness field, were as follows:

1) The 500 mb height field ϕ^τ was smoothed according to the equation

$$\phi^{(N)\tau} = \phi^\tau + \frac{1}{4}\nabla^2\phi^\tau \qquad (32)$$

The grid size used when calculating the finite difference Laplacian was 600 km.

2) The value of $\xi^{\tau+1}$ was calculated from the equations

$$\left.\begin{aligned}\left(\frac{\partial\xi}{\partial t}\right)^\tau &= \frac{m^2}{f}J\left(\xi^\tau, \phi^{(N)\tau}\right)\\ \xi^{\tau+1} &= \xi^{\tau-1} + 2\Delta t\left(\frac{\partial\xi}{\partial t}\right)^\tau\end{aligned}\right\} \qquad (33)$$

where Δt is the length of the time step, taken 1 hour during all forecasts discussed here.

It should be emphasized that the $\phi^{(N)\tau}$ field was used only to advect ξ, while the barotropic forecast was made in the usual manner.

3) When a forecast for ϕ_T was needed, after τ_p time steps, it was obtained by solving the equation

$$\frac{1}{2}\phi_T^{\tau_p} - \frac{1}{8}\nabla^2\phi_T^{\tau_p} = \xi^{\tau_p} \qquad (34)$$

for $\phi_T^{\tau_p}$, by a Liebmann process. The convergence is very rapid. A satisfactory solution for a 24-hour forecast was usually obtained after 7—8 iterations. A 600 km grid size was used during the solution, for the same reason as when the initial value of ξ was calculated. All boundary values were kept constant during the computations.

The use of $\phi_T = \phi_{500} - \phi_{1000}$ instead of $\phi_T = \frac{1}{2}(\phi_{z=h} - \phi_{z=-h})$ involves the assumption that the height of the atmosphere, and in particular, the height where the boundary condition $w = 0$ may be applied, is twice the height of the 500 mb surface. This puts the height of the atmosphere approximately at tropopause level.

The time needed for the computation of a combined 24-hour thickness and barotropic forecast with a time step of one hour is about 45 minutes or approximately 75 % more than a conventional barotropic forecast. The time for program read in, data checking and printing of the results is included in this figure.

3. Discussion of the results

Analyses for the following days were chosen as a basis for the forecast study:

0300 GMT, 26th 27th and 28th November, 1951.

0300 and 1500 GMT 30th January and 0300 GMT 31st January, 1956.

All analyses were made by the conventional method, except the 500 mb analysis 1500 GMT January 30th, which was made using the numerical method developed by Bergthorsson and Döös (1955). Forecasts for 24-hours were made for all the cases.

The synoptic situations in the two series are rather different. The one from 1951 is characterized by strong zonal flow, with fast moving and rapidly developing depressions, especially in the western and northern parts of the North Atlantic Ocean. Outflow of

Table VI

Date and Time	24-hour forecast					48-hour forecast				
	r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x
	Small region									
26/11-51 03	0.91	115	105	50	0.44	—	—	—	—	—
27/11-51 03	0.89	94	125	61	0.65	0.71	113	178	121	1.07
28/11-51 03	0.66	117	101	84	0.72	—	—	—	—	—
30/1 -56 03	0.65	73	50	49	0.67	0.66	146	113	105	0.72
30/1 -56 15	0.86	82	70	39	0.47	—	—	—	—	—
31/1 -56 03	0.78	111	75	74	0.66	—	—	—	—	—
Mean	0.79	99	88	60	0.60	—	—	—	—	—
	Large region									
26/11-51 03	0.86	110	139	72	0.65	—	—	—	—	—
27/11-51 03	0.79	105	110	74	0.70	0.57	120	156	134	1.12
28/11-51 03	0.76	105	124	80	0.76	—	—	—	—	—
30/1 -56 03	0.72	79	60	52	0.66	0.32	126	126	140	1.11
30/1 -56 15	0.62	83	86	71	0.85	—	—	—	—	—
31/1 -56 03	0.67	98	111	87	0.88	—	—	—	—	—
Mean	0.74	97	105	73	0.75	—	—	—	—	—

cold continental air over the ocean, in the rear of the depressions is very marked in these regions, especially on the 27th and 28th.

The series from 1956 is characterized by a persistent cold high centered over Scandinavia. The air over the North Atlantic is relatively warm. Cold air outflow from the North-American continent occurred just before the beginning of the series. In the latter part of the period a cold air advection occurred only in the region around southern Greenland.

Correlation coefficients and other quantities indicating the quality of the thickness forecasts are given in table VI. The meaning of the various symbols is the same as in table II and III, regions for which the correlations were made are indicated on fig. 2.

Little can be concluded from only six cases, but a comparison between the results of the thickness forecasts with results of the barotropic forecasts presented by Döös (1955) and those in this article, seem to indicate that the accuracy of the 24-hour thickness forecasts in general is somewhat less than of the 24-hour barotropic forecasts.

The case from 0300, 27th November, 1951 may be taken as one which fairly well illustrates both the success and shortcomings of the method. During the period a deep surface low near Nova Scotia moved northeast towards southern Greenland, and the associated trough at the 500 mb level deepened considerably, forming a closed circulation. At the beginning of the period the thickness lines were running relatively straight from North America across the ocean towards Europe, except for a small disturbance in eastern United States and Canada. After 24 hours the disturbance in the thickness lines was very pronounced. Warm air had been advected towards Greenland and cold air towards Newfoundland and the eastern seaboard.

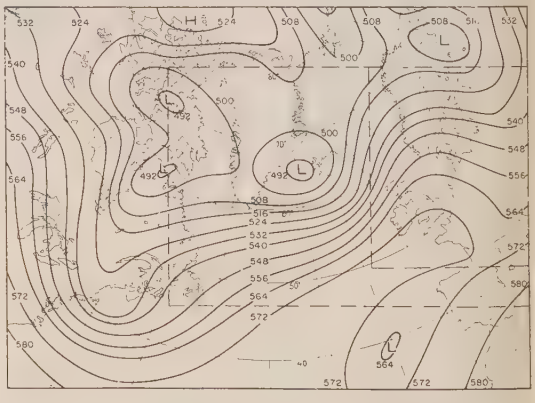


Fig. 2. 500 mb map 27. 11. 1951 03 GMT. The areas, for which correlation coefficients for thickness forecasts were calculated, are indicated by dashed lines.

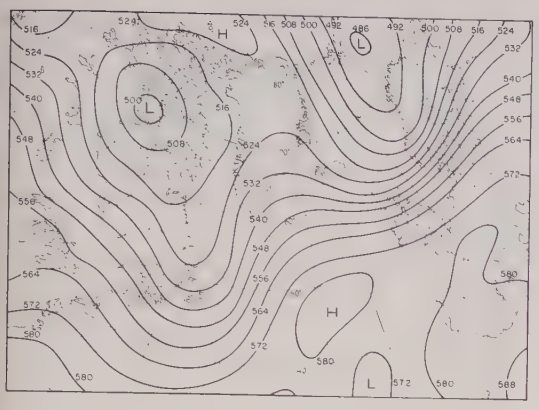


Fig. 3. 24 hour forecast map of 500 mb contour field, based on analysis at 27. 11. 1951 03 GMT.

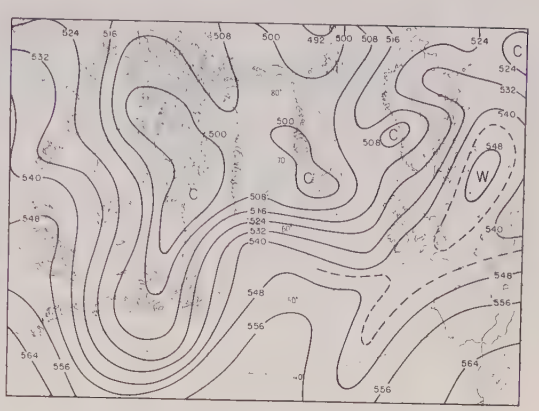


Fig. 5. 24 hour forecast for relative topography between 500 and 1000 mb surfaces, based on analyses at 27. 11. 1951 03 GMT.

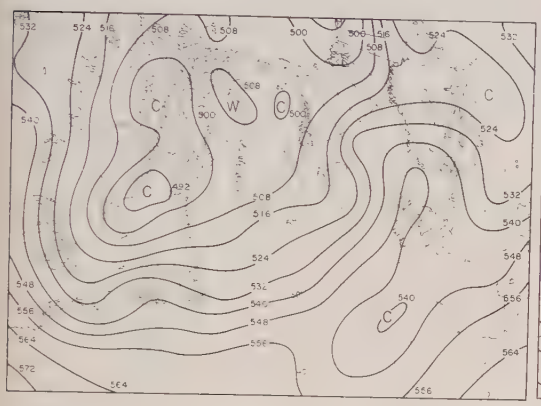


Fig. 4. Analysis of relative topography between 500 and 1000 mb surfaces, 27. 11. 1951 03 GMT.

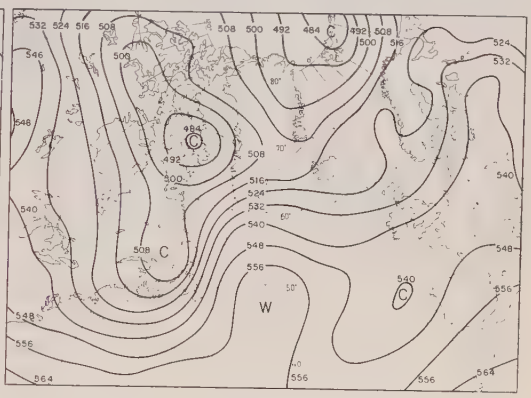


Fig. 6. Analysis of relative topography, 28. 11. 1951 03 GMT.

Cold air advection had occurred over the ocean around Iceland and towards Scandinavia behind a deep complex low moving eastwards, while over Europe the thickness troughs and ridges moved rather steadily eastwards.

All these features are apparent in the thickness forecast, and a change in the relative topography south of Greenland amounting to around 250 meters was forecast with an accuracy of about 30—40 meters. The cold outbreaks are greatly exaggerated, however, both in the areas south of Newfoundland and around Iceland, while over eastern Canada and United States the forecast produced too large thickness values. The errors tend in general to exaggerate the thickness gradient between the warm and the cold air, causing fictitious "frontogenesis". This is in itself an interesting fact.

One might suspect that the following sources of errors are the most important ones:

a) Differences between the forecast and actual wind fields at 500 mb, which will influence the advection.

b) The assumption of constant stability in time and space.

c) The adiabatic assumption.

The barotropic forecasts were relatively good for all cases in both series. On the whole the barotropic forecast tended to give gradients which were too weak. One might therefore expect that the forecast would give too short displacements, but as already indicated, the results showed an error in the other direction. Significant errors in the barotropic forecasts were observed in limited

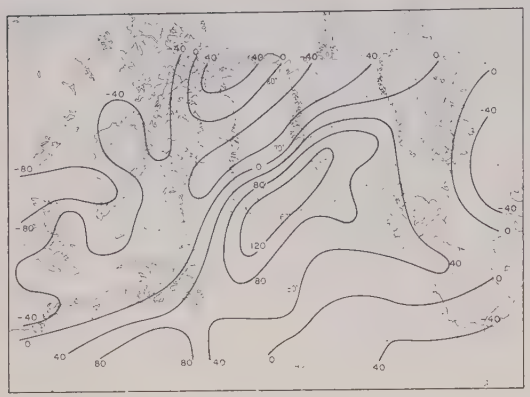


Fig. 7. Mean error map for the 24 hour relative topography forecasts in the series 0300 GMT 26th to 28th November 1951. The error is calculated as the quantity (gpm) to be added to the prognotic thickness values in order to get the observed ones.

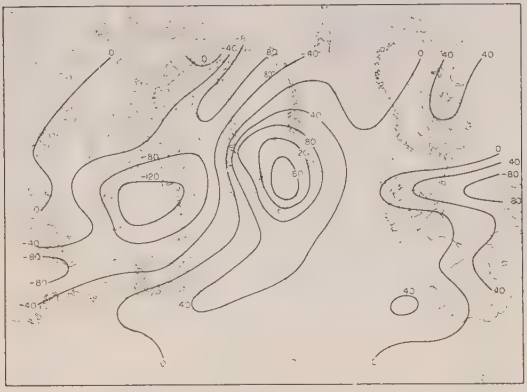


Fig. 8. Mean error map for 1956 series.

Fig. 9. Mean error map for both series.

areas, however, and some of the errors of the thickness forecasts could be explained by this fact.

In order to check the assumption that $k_N = \frac{1}{2}$ its values were calculated at 39 stations at 0300 GMT 27th November, 1951. The mean of these values was 0.53, but individual values varied from 0.26—0.73, one station (Tampa, Florida) reaching 0.8. Values considerably lower or higher than the average seemed to be representative for significant areas of the map. A 24 hour forecast from the time just mentioned was therefore made with $k_N = 0.6$. The effects on the forecast were negligible in all regions, and no further experiments along this line were therefore made.

Errors due to the adiabatic assumption are in individual cases masked by others, but on mean error maps one should expect that spurious errors, such as those due to the faults of the barotropic forecast to be smoothed considerably. The mean error maps for the separate series from 1951 and 1956, and also for all six cases treated are presented as fig. 7, 8 and 9. A feature already apparent on the error maps for the individual cases stands out clearly on the mean error maps. The forecast gives too high values for the thickness over the continents, particularly the North American one, and too low over the North Atlantic Ocean. On figure 9 the line of zero average errors follows closely the east coast of North America, with the axis of maximum positive errors lying parallel to it about 700—800 km off shore. The error maps for the separate series show similar main features. The series from November 1951, which had strong zonal circulation and relatively large outflow of cold air from the North American continent over the North Atlantic, also has larger positive errors over the ocean than the series from 1956, when the zonal circulation was weaker. The largest positive errors in the 1956 series occur near southern Greenland, which was also the area where cold air outflow was strongest.

Separation of errors due to changing stability from those due to the adiabatic assumptions might be difficult, since low stability is generally associated with heat sources and vice versa.

The geographical distribution of the mean errors and their variation with the air flow at

1,000 mb which was described above, suggest, however, that non-adiabatic effects contribute greatly to the errors of the thickness forecast, heat sources being mainly associated with the oceans, and heat sinks with the continents. If we for a moment assume that the mean errors are due only to non-adiabatic effects, and further that the value of c_p may be used for the specific heat of the 500—1,000 mb layer, we find that an error of 100 m corresponds approximately to a heat flow of 600 ly/24 hours to or from the layer. The mean error for 28 grid points for which data were available in North America, is 40 m, giving a net heat loss of 240 ly/24 hours for the layer in question. A corresponding value for 41 points over the ocean is a net gain of 170 ly/24 hours.

It may be of interest to compare these figures with values given by those who have investigated the heat budget of the atmosphere and the oceans, in order to see whether the mean errors attributed here to the non-adiabatic assumption, are of a similar order of magnitude as generally accepted values for the heat exchange.

LONDON (1952) gives 200 ly/24 hours as an average value for the radiation losses of the entire troposhere of the northern hemisphere in March. From London's diagrams of the radiative cooling, and assuming that the 1,000—500 mb layer contains about 2/3 of the mass of the troposphere, it may be estimated that roughly 75 %, or 150 ly/24 hours is taken from this layer. If this is the case, the value given by the mean forecast error is of the same order of magnitude. Taking again 150 ly/24 hours as the radiative cooling for the 1,000—500 mb layer over the ocean and combining it with the net heat gain figure of 170 ly/hour estimated for the layer over ocean from the thickness error, one gets that the heat supply from the ocean to the layer should be about 320 ly/hour. This value is of comparable magnitude with the values for total amount of energy given off in winter from the North Atlantic Ocean, according to SVERDRUP (1945).

In order to obtain a better picture of how the heating and cooling affects the forecast one may refer to the thermodynamic equation, which was one of the equations used by Fjörtoft to obtain the forecasting equation (30) for ψ_T, and include in it a term $\frac{dQ}{dt}$ expressing addition or subtraction of heat; the thermodynamic equation (12) becomes:

$$\frac{\partial \ln \Theta}{\partial t} = -\mathbf{v} \cdot \nabla \ln \Theta - ws + \frac{dQ}{dt} \qquad (35)$$

During the integration of the equations with $\frac{dQ}{dt}$ included, it is convenient computationally to assume a certain distribution of $\frac{dQ}{dt}$ with height. The distribution assumed here is exponential,

$$\frac{dQ}{dt} = \left(\frac{dQ}{dt}\right)_G e^{-\gamma\left(\frac{z}{h}+1\right)} \qquad (36)$$

where $\left(\frac{dQ}{dt}\right)_G$ is the heat added at the ground (which can of course be negative), γ is a constant, and h represent half the thickness of the atmosphere, z is the vertical coordinate, its zero is taken at the middle of the layer, z is therefore $= -h$ at the bottom of the layer and $+h$ at the top.

Following through the steps to obtain the forecast equation one gets instead of equation (30)

$$\frac{D^{(N)}}{Dt}\left(\psi_T - \frac{1}{2}\psi_T^{(N)}\right) = \\ = \Gamma\left[\left(\frac{dQ}{dt}\right)_G - \left(\frac{dQ}{dt}\right)_T\right] \qquad (37)$$

where Γ is a constant depending on γ, and $\left(\frac{dQ}{dt}\right)_T = \left(\frac{dQ}{dt}\right)_G e^{-2\gamma}$. It can be seen from equation (37) that an error arising from neglecting a heat source (positive or negative) would lead to an error in the thickness forecast, which would be carried on with the wind. If the heat source persisted, the distribution of the error would form a trail downwind, and if the heat source was an area over which the air streamed, instead of a point source, the error would be cumulative. It is therefore natural, considering the flow nearest the surface in the test cases, that the maximum positive error should form a belt along the coast of North America, which then stretches

across the ocean from Greenland to North Western Europe. This would also explain the rather peculiar error distribution over south and southeast Greenland for the 1956 series. The air after being warmed over the ocean and causing the large positive errors, then flows towards the Greenland glacier and is cooled again gradually, but not rapidly enough to neutralize the previous heating immediately. Thus we have positive errors over parts of the glacier, which, however, decrease gradually downstream.

If one would try to correct the forecast by using eq. (37) instead of (30) as a forecast equation, some scheme would have to be found to calculate $\left(\frac{dQ}{dt}\right)$, $\left(\frac{dQ}{dt}\right)_T$ could probably be neglected in comparison to $\left(\frac{dQ}{dt}\right)_G$ rather than using its assumed dependence on $\left(\frac{dQ}{dt}\right)_G$. The method of determining $\left(\frac{dQ}{dt}\right)_G$ is too complicated to discuss in any detail here, but it might be suggested that an expression similar to the one proposed by MINTZ, (1955) might be used with advantage.

The results just described indicate that the main systematic errors in the forecasts are due to the adiabatic assumption and are closely connected with flow at low levels. These results suggest, that the method could be used to an advantage in forecasting thickness in layers that do not extend to the ground, but whose lower boundary is, for example, the 850 mb surface. The upper boundary need not be the 500 mb level, but might just as well be the 300 or 250 mb level.

Beside the forecasting procedure based on equation (30) which has already been described, experiments were also made to use equation (29) and use either entirely unsmoothed ϕ fields, or use various degrees of only initial smoothing of the ϕ field. 24-hour thickness forecasts were made with data from the November 1951 series with this method. The results were not encouraging. The displacements of the thickness lines became on the whole too large, a large number of wiggles with excessive amplitudes were introduced on the thickness lines and the thickness gradient in middle latitudes became much too large. Equation (30) clearly showed its superiority over (29) as a forecasting tool.

Acknowledgement

The authors wish to thank Dr. Bert Bolin, who suggested this work, for his great interest and encouragement and for reading the manuscript. We are further indebted to the staff of the International Meteorological Institute in Stockholm for many fruitful discussions. Finally we wish to thank Mrs. B. Kreiss and Mr. A. Bring for their great help in coding the computations, Mrs. I. Hedberg who prepared the tapes, and Mrs. B. Dahlborg who carried out the drafting.

REFERENCES

BERGTHORSSON, P., and DÖÖS, B. R., 1955: Numerical Weather Map Analysis. *Tellus* **7**, 5.

DÖÖS, B. R., 1956: Automation of 500 mb Forecasts through Successive Numerical Map Analyses. *Tellus* **8**, 1.

FJÖRTOFT, R., 1952: On a Numerical Method of Integrating the Barotropic Vorticity Equation. *Tellus* **4**, 3.

— 1955: On the Use of Space-Smoothing in Physical Weather Forecasting. *Tellus* **7**, 4.

LONDON, J., 1952: The Distribution of Radiational Temperature Change in the Northern Hemisphere during March. *J. Meteorology* **9**, 145—151.

MINTZ, Y., 1955: On Heating Models for Extended and Long-range Numerical Weather Forecasting and the Study of the General Circulation. *Scientific Report No. 2, University of California*, 1955.

STAFF MEMBERS, INSTITUTE OF METEOROLOGY, UNIVERSITY OF STOCKHOLM, 1954: Results of Forecasting with the Barotropic Model on an Electronic Computer (BESK). *Tellus* **6**, 2.

SVERDRUP, H. U., 1942: Oceanography for Meteorologists, New York, 1943, p. 229—231.

A Note on Verification of Prognostic Charts

By SVERRE PETTERSSEN, University of Chicago[1]

(Manuscript received January 2, 1957)

1. Correlations of pressure changes

In recent forecasting experiments it has been customary to verify the prognostic charts by computing the correlation coefficients between the observed and predicted pressure changes. The justification has been that correlations between predicted and observed pressures would not correctly reflect skill, since in many cases high correlations would result from persistence.

While correlations of observed and predicted changes certainly remove similarities resulting from persistence they also tend to remove the measure of similarities resulting from travel of pressure systems. To illustrate, it suffices to consider a sinusoidal pressure wave with length L, amplitude A, and phase velocity C which travels along an axis x. The equation for the wave profile may then be written

$$Y_0 = A \sin \frac{2\pi}{L}(x - Ct) + N_0 \qquad (1)$$

where N_0 symbolizes some kind of noise superimposed upon the primary wave. The noise may be real or due to errors in observations and the analyses. In any case, it will be assumed that the maximum value of N_0 is very much smaller than A. In general, the forecasting problem will involve changes in A, L and C, but for the purpose of illustration it will be assumed that the wave is permanent, in the sense that its speed and shape are maintained. Next, let it be assumed that the movement of the primary wave was predicted perfectly, so that at the end of the forecast period the prediction was

$$Y_p = A \sin \frac{2\pi}{L}(x - Ct) + N_p$$

where N_p signifies some kind of noise ($A > N_p$) which was not predicted.

Now if the forecast period were such that $Ct = \frac{1}{2} L$, the observed and predicted changes would be

$$\Delta Y_0 = -2A \sin \frac{2\pi}{L} x + \Delta N_0,$$

$$\Delta Y_p = -2A \sin \frac{2\pi}{L} x + \Delta N_p$$

respectively.

Since the extreme values of N were supposed to be very much smaller than A, it is evident that a correlation close to unity would be found. On the other hand, if the forecast period were such that $Ct = L$, the changes would be

$$\Delta Y_0 = \Delta N_0 \quad \Delta Y_p = \Delta N_p$$

and the correlation would be determined exclusively by the noise.

In practice this would mean that correlation of pressure (or height) changes would

[1]) The research reported in this note has been made possible through support and sponsorship extended by the Geophysics Research Directorate, Air Force Cambridge Research Center, under Contract No. AF 19(604)—1293.

give high credit to forecasts in cases where the final state tended to be out of phase with the initial state, and it would discredit forecasts in which in-phase relationships were involved.

It can readily be shown that verification by pressure (or height) changes tend to give undue credit to poor forecasts in cases where the distance traveled by the system is small in comparison with its dimensions. To illustrate, we return to the sinusoidal wave. Leaving the noise out of consideration, it is readily found that the correlation coefficient $R(\Delta)$ between the observed and predicted changes is expressed by

$$R(\Delta) = \pm \cos \pi \frac{D_0 - D_p}{L} \quad (D_0 D_p \gtrless 0) \quad (2)$$

Here D_0 and D_p signify the observed and the predicted displacements, and plus or minus are chosen according as D_0 and D_p have the same or opposite sign.

Since $D_0 - D_p$ is the error in the forecast displacement, it will be seen that $R(\Delta)$ remains positive as long as the error remains less than one-half of the wavelength.

In connection with the forecasting of upper waves it would seem reasonable to consider a forecast to be useful if the error in the predicted displacement were less than the displacement itself. As an example, let us consider an average upper wave of length 60° latitude and a 24-hour displacement of 10°. With the above definition of usefulness one would find that only correlation coefficients in excess of 0.87 would indicate success. On the other hand, if the wavelength were 30° and the displacement 15°, any positive value of the correlation coefficient would signify success. From the simple cases discussed above it would appear that the interpretation of correlation of changes, in terms of usefulness, is rather complex.

2. Correlation of pressures

Since the main purpose of prognostic charts is to provide a basis for the issuance of specific forecasts, and since the circulation patterns, rather than the change patterns, at the end of the forecast period are important, it appears that the usefulness of prognostic charts can more readily be assessed from correlations (or similar measures) of the agreement between the predicted and the observed pressure distributions. Since the correlations reflect only proportionately, the root-mean-square error, or some similar measure, would supply additional information. Whatever system of verification is used, the zero-level (determined by the degree of persistence) must be considered.

As regards real persistence it must be admitted that considerable skill on the part of the forecaster is required to predict with success that the changes throughout an appreciable period will be insignificant. If, in such a case, an individual prognostic chart verifies no better than straight persistence, a positive measure of skill would seem in order. It appears, therefore, that there can be but little justification for rating *individual* forecasts on a scale in which the persistence pertaining to the *case itself* marks the zero level. Any such scheme, as well as verification by pressure changes, will emphasize the forecasting of out-of-phase relationships and penalize forecasts of in-phase relationships, whether these latter are due to real persistence or accidental similarity due to travel of systems.

A measure of the skill in which similarity between the initial and final chart is accounted for can be justified for a *representative sample* of forecasts, but the purpose of such a score would be to measure the yield of the effort as compared with no effort at all. To illustrate, let it be assumed that no effort is made to predict the future state and that, instead, the actual chart is presented as a prognostic chart for the end of, say, a 24-hour period. Over a representative period a certain correlation would be found. Whether this correlation resulted from real persistence or accidental similarities is immaterial since, obviously, the procedure involved no skill.

Let $R(IV)$ and $R(PV)$ denote the correlation coefficients between the initial and verifying charts and between the prognostic and verifying charts, respectively, both pertaining to a representative sample of forecasts. Obviously, the forecasts would reflect some skill if $R(PV) - R(IV) > 0$. The maximum score obtainable (for perfect forecasts) would be $1 - R(IV)$. A correlation skill score S, similar to that derived from contingency considerations, could then be defined as

$$S = \frac{R(PV) - R(IV)}{1 - R(IV)} \quad (3)$$

The score would be zero if current information were presented as forecasts. A positive score for a representative sample of forecasts would indicate some degree of skill, and different forecasting techniques, applied to the same sample of cases, can be compared by the scores obtained.

In a recent forecast experiment (PETTERSSEN, ESTOQUE and HUGHES, 1956) it was found that $R(IV)$ for 24-hour sea-level forecasts for the United States in winter was about 0.6. For such low values a relatively short period suffices for determining whether or not a forecasting technique is capable of yielding useful results. In the same experiment it was found that $R(IV)$ at the 500 mb level was about 0.9. For such high values of $R(IV)$ a lengthy trial would be required to obtain significant results. Obviously, as $R(IV)$ approaches unity, prediction ceases to be a problem.

3. A scheme of verification

In the aforementioned forecast experiments the following verification measures were used

$R(PV)$—the correlation coefficient between prognosticated and verifying pressures (or heights).

$R(IV)$—the correlation coefficient between the initial and the verifying pressures (or heights).

S —the correlation skill score as defined by Eq. (3).

$R(\Delta)$ —the correlation coefficient between the predicted and observed changes.

E —the root-mean-square error of the predicted changes.

C —the root-mean-square of the observed changes.

N —the number of occasions when $R(PV)$ exceeded $R(IV)$.

M —the number of occasions when E was less than C.

While no single verification measure will describe the efficacy of a forecasting technique, it was found that the measures S, E, C, N and M, used in combination, are adequate for comparing different methods. Examples of such comparison will be given elsewhere (PETTERSSEN, 1957). Here, it may be noted that not only the skill score but also the value of E, in comparison to that of C, has been found useful. In comparing different methods not only the question of average performance but also that of stability has to be considered, and the measures M and N have been found useful for assessing the stability.

4. Conclusion

Forecasting research has now progressed to a stage where it has become necessary to compare different prediction models, techniques and procedures. While the verification scheme outlined here may not be ideal, it is hoped that it will provoke discussion, and result in the establishment of uniform methods. A yardstick for comparison of prognostic techniques is much in need.

REFERENCES

PETTERSSEN, S., ESTOQUE, M. A., and HUGHES, L. A., 1956: An experiment in prognostication. Scientific Report No. 12, Contract AF 19(604)—1293. Department of Meteorology, University of Chicago.

PETTERSSEN, S., 1957: A review and summary of recent forecasting research. Final Report, Contract AF 19(604)—1293. Department of Meteorology, University of Chicago.

Preliminary Research in Objective Analysis

By D. H. JOHNSON, Central Forecasting Office, Dunstable

(Manuscript received November 13, 1956, revised March 8, 1957)

Abstract

Simple mathematical surfaces are used to interpolate numerically within two-dimensional networks of 500 mb contour heights and winds. The effect on the errors of interpolation of varying (i) the relative weights given to the measured contour heights and winds, (ii) the complexity of the mathematical surface, and (iii) the layout and dimensions of the network is examined. It is concluded that in areas where observations are reasonably dense, numerical analysis will be at least as satisfactory as the currently used subjective graphical method.

Introduction

The potential merits of objective analysis as a preliminary to numerical forecasting are well-recognised and objective wind interpolation may find application in other synoptic work.

Studies of objective analysis have been made in the United States (PANOFSKY 1949, GILCHRIST and CRESSMAN 1954) and in Sweden (BERGTHORSSON and DÖÖS 1955). The results of the first phase of parallel work which is in progress at Dunstable, are given below.

Method of Approach

As a first step, attention has been confined to the basic problem of interpolating between a two-dimensional network of observations without regard to conditions at other levels or at other times, although it seems probable that fully successful schemes of objective analysis will eventually take into account continuity in time and consistency in the vertical.

One method of interpolating for the height of a contour surface at a given grid point is to represent the contour height over some area surrounding the grid point mathematically as a function of position, the function being chosen so as to give the appropriate values at the observing stations. By applying the geostrophic relation, the observed winds can also be taken into account.

Before the analysis can serve as a basis for a numerical forecast, it may be desirable to smooth the wind or contour fields so as to leave only those scales of disturbance which are to be forecast and which are consistent with the forecast equations. However, the proper degree of smoothing will depend upon the nature of the forecast and the particular mathematical model selected. So that the objective analysis should have general application, the principal aim in the work presently described has been to get accurate values for interpolated winds and heights. Smoothing to eliminate unwanted, though real, wavelengths has been taken to fall within the province of the forecast rather than that of the analysis.

Attempts were made to represent the contour height, H, as quadratic and cubic functions of suitable co-ordinates on the earth's surface. The co-ordinates chosen correspond to rectangular cartesian co-ordinates on a stereographic projection of the earth's surface from the south pole onto the tangent plane

through the north pole. For the quadratic, H is expressible in the form

$$H = ax^2 + 2hxy + by^2 + 2gx + 2fy + c,$$

where a, h, b, g, f and c are constants determined from the observations. Because the observations contain errors, it is advisable not to fit them exactly but to introduce a degree of smoothing using the principle of least squares. A least-squares solution is obtained by choosing a, h, etc., so as to minimise the expression

$$\sum_i (H - H_i)^2 + T^2 \sum_i (\mathbf{V} - \mathbf{V}_i)^2, \qquad (1)$$

where the H_i are observed contour heights, the $\mathbf{V}_i$ are observed winds, $\mathbf{V}$ is the geostrophic wind, T decides the relative weight to be given to contour height and wind observations, and the summation is taken over i observing stations. Components of $\mathbf{V}$, V_x and V_y, are given by $-\frac{\beta g}{l}\frac{\partial H}{\partial y}$ and $\frac{\beta g}{l}\frac{\partial H}{\partial x}$ respectively, where g, l have their usual significance and β is a magnification factor arising from the choice of projection. Equation (1) is minimised when a, h, etc., satisfy the six simultaneous linear equations obtained by equating to zero the expressions given on partial differentiation of expression (1) by a, h, etc., in turn.

Choice of weight

The weighting factor T, which occurs in expression (1), holds a balance between contour heights and winds and depends upon, amongst other things, the units of H and $\mathbf{V}$. The terms $\sum_i (H - H_i)^2$ and $T^2 \sum_i (\mathbf{V} - \mathbf{V}_i)^2$ are thereby permitted to attain the same order of magnitude; if these terms differed greatly, the least squares process would ignore the smaller of the two. In the computations described below it was convenient to work with $\mathbf{V}$ in ft/sec and H in ft. For the two terms of (1) to have equivalent dimensions, the units of T have been taken as seconds. The value of $\sum_i (H - H_i)^2$ depends on the random errors of observation of contour height and also upon a kind of topographical error which arises because quadratic surfaces comprise a restricted class of shapes; the true contour surface will not in general coincide with one of these shapes, although there may be a sufficient resemblance over a limited area. $\sum_i (\mathbf{V} - \mathbf{V}_i)^2$ contains ageostrophic effects as well as errors of observation and errors of a topographical character. The topographical errors will vary according to the function used to represent the contour surface, so T itself may require adjustment according to the function chosen.

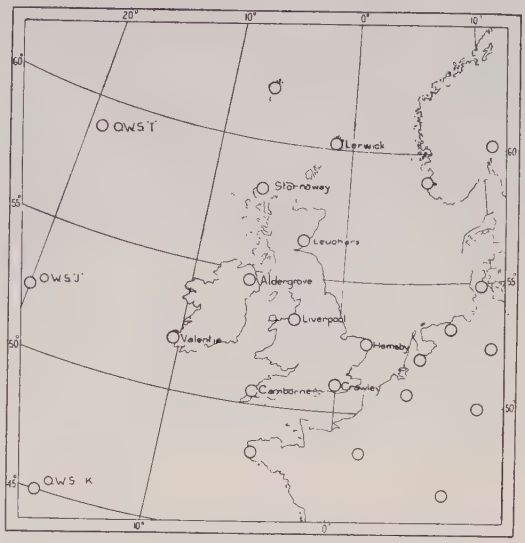


Fig. 1. Key to stations used in the computations.

Computation No. 1

To determine an optimum weighting factor a quadratic surface was fitted to 500 mb heights and winds from six British stations, Aldergrove, Camborne, Crawley, Hemsby, Leuchars and Valentia, for 100 occasions selected from the first nine months of 1954. Computations were made in each case for the ten integral values of T from 0 to 9. Cases were chosen when there were reports of contour height and wind from all six stations, but otherwise dates were picked at random. To test the value of the surfaces for interpolation, the implied Liverpool heights and geostrophic winds were compared with Liverpool upper air observations. Figure 1 shows the disposition of the stations used for these and subsequent calculations. The root-mean-square differences between observed and computed contour heights and winds at Liverpool for the 100

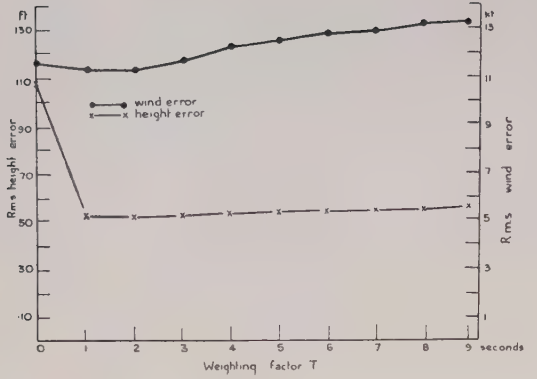


Fig. 2. Computation No. 1. Errors of representation at Liverpool.

cases are given by Figure 2. Values for $T = 0$ were special since, if the winds are ignored, the six contour heights can be fitted exactly by a quadratic and there is no smoothing. The lowest interpolation errors occurred in the range $T = 1$ to $T = 2$, where the r.m.s. wind difference was 11.4 kt and the height difference was 52 ft.

The result of this experiment is comparable with that obtained by GILCHRIST and CRESSMANN (1954), who computed an objective analysis for a single occasion over a grid of 19 × 19 points covering most of North America and a part of the Atlantic. Their method was very similar to that described above: a quadratic surface was fitted to observations within a limited distance of a grid point to get an interpolated contour height. The interpolated heights were compared, not with observations, but with the mean of three subjective analyses. The minimum height "errors" occurred with a weighting factor corresponding to $T = 2$.

Figure 3 shows the variation with T of the r.m.s. height and wind differences when summed over the six stations to which surfaces were fitted and over the 100 cases. In the range $T = 1$ to $T = 2$ the r.m.s. height difference rose from 21 to 30 ft. whilst the r.m.s. wind difference fell from 15 to 12 kt (25 to 20 ft/sec). So the two terms of expression (1) did in practice attain the same order of magnitude: — at $T = 1$, $\Sigma_i (H - H_i)^2 = 400$ ft^2 while $T^2 \Sigma_i (\mathbf{V} - \mathbf{V}_i)^2 = 625$ ft^2 T^2 having units of sec^2;

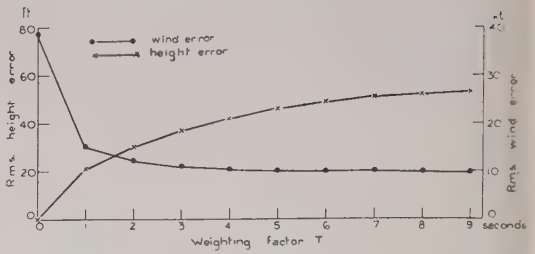


Fig. 3. Computation No. 1. Errors of representation at the six stations.

at $T = 2$, $\Sigma_i (H - H_i)^2 = 900$ ft^2 while $T^2 \Sigma_i (\mathbf{V} - \mathbf{V}_i)^2 = 1{,}600$ ft^2.

Computation No. 2

It was remarked above that the results obtained for $T = 0$ in Computation No. 1 were special since the contour heights at the six stations were exactly fitted. The computations were repeated for 40 of the 100 cases using observations from Stornoway and Lerwick in addition to the six previously used. In this case, even at $T = 0$, there were more pieces of information than were required to determine the co-efficients of the quadratic uniquely.

In Figure 4 the r.m.s. height and wind differences at Liverpool for the eight station computations are comperad with results for

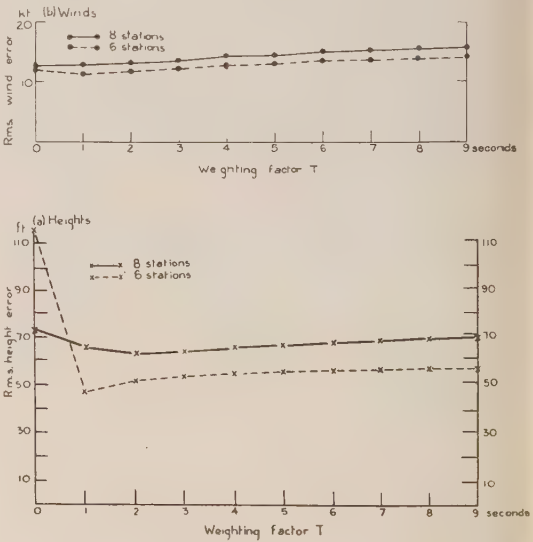


Fig. 4. Computation No. 2. Errors of representation at Liverpool.

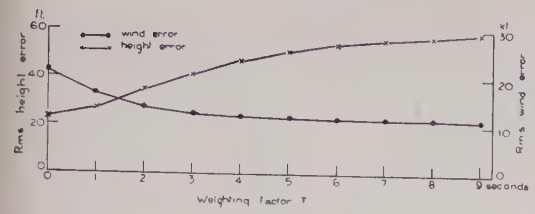


Fig. 5. Computation No. 2. Errors of representation at the eight stations.

the same 40 cases from Computation No. 1. Except for the heights at $T = 0$, the Computation No. 2 differences exceed those of Computation No. 1. The inclusion of the two extra heights reduced the r.m.s. Liverpool height difference considerably. The improvement here was due entirely to the beneficial effect of smoothing out errors of observation, etc., since results for $T = 1$ to $T = 9$ suggest that it was generally undesirable to extend the area over which the quadratic was fitted beyond that covered in Computation No. 1. As in Computation No. 1, values of T in the range from $T = 1$ to $T = 2$ appeared to be satisfactory from the point of view of minimising the errors of representation. Neither the wind nor the height errors were very sensitive to variations in T.

Figure 5 gives the standard errors of fitting at the eight stations. These may be compared with the corresponding values for Computation No. 1 in Figure 3. Computation No. 1 values were smaller except for the winds at $T = 0$.

Computation No. 3

To examine whether a surface of greater complexity than the quadratic was of more value for interpolation, a cubic surface was fitted to the six stations used for Computation No. 1, in 38 cases. Computations were made for six values of T. $T = 1$, $2^{\frac{1}{2}}$, 2, $2^{\frac{3}{2}}$, 4 and 6.

The r.m.s. differences between observed and interpolated heights and winds at Liverpool are given in Figure 6. Neither the height nor the wind differences showed much variation over the range of T explored, but such variations as there were followed a different pattern from the quadratic results of Computation No. 1. The lowest Liverpool wind difference occurred at about $T = 4$ whereas the height

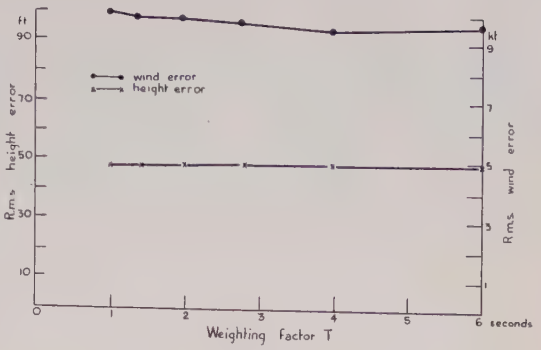


Fig. 6. Computation No. 3. Errors of representation at Liverpool.

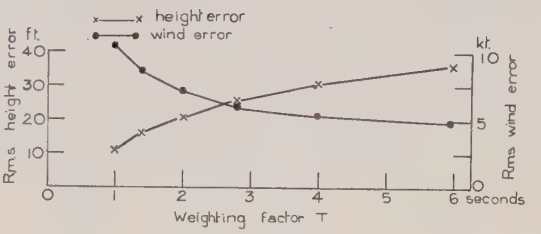


Fig. 7. Computation No. 3. Errors of representation at the six stations.

errors increased with T throughout the range examined. The cubic results are best compared with the quadratic six-station values given in Figure 4 which, with two exceptions, refer to the same occasions. The two exceptions do not materially affect the comparison. The cubic surfaces gave a slightly better fit than the quadratic at Liverpool.

Figure 7 gives the r.m.s. errors of representation at the six stations whose observation were used to obtain the cubics. The values are naturally lower than those obtained for the quadratic (figure 3) since the cubics have a greater degree of freedom.

The cubic results show how the choice of weighting factor T may depend upon the purpose for which the interpolation is required. A preliminary to the preparation of a numerical forecast is the evaluation of the Laplacian derivative of the initial contour field. Normally, the Laplacian is derived from the set of interpolated contour heights by means of the finite difference approximation. If the contour field is represented by some mathematical function, it is possible to obtain the Laplacian in a second way, namely by

direct differentiation of the function itself. For this latter purpose it is desirable that the function should not just give an accurate interpolation for height at the grid point, but should also approximate closely to the true contour surface over a certain area surrounding the grid point. In Computation No. 3 the contour height interpolation error was lowest at $T = 1$ (Figure 6), but Figure 7 shows that at $T = 1$ the heights at the six stations were fitted with an error of only 10 ft. This closeness of fit was unrealistic since the standard error of measurement of 500 mb contour height at British stations is 30 ft. Futhermore the winds at the six stations were fitted at $T = 1$ with an error of about 10 kt, which is larger than the standard error, 3 kt, of British 500 mb wind observations. At $T = 4$, on the other hand, the r.m.s. height error at the six stations was 30 ft and the r.m.s. wind error was 5 kt. Thus, if a realistic fit is required over an area of 300 mi or so centred upon Liverpool, a weighting factor $T = 4$ is to be preferred to $T = 1$.

According to MURRAY (1954) the r.m.s. difference between observed winds and geostrophic winds measured from the C.F.O. working charts is 14 kt at 500 mb over the British Isles. It is interesting to note that the wind values interpolated from the cubic surfaces when compared with independent observed winds gave a r.m.s. difference of less than 10 kt. Murray shows that the 14 kt can be broken down into the following components: 9 kt due to the ageostrophic wind; 4 kt due to turbulent wind fluctuations; 3 kt due to errors in the observed wind; 9 kt due to errors inherent in the graphical technique. The error of representation of the wind field given by the true contour pattern will contain the first three of these four components and should therefore have a root-mean-square value of about 10 kt. On this basis it appears that the 5 kt error of representation of the six stations, given by the cubics for a weighting factor $T = 4$, is lower than is justified. However, in many techniques of numerical forecasting the contour height field is used essentially as a stream function and a closer fitting of the winds than seems proper on purely geostrophic grounds may give a better approximation to such a stream function than is otherwise obtained.

PANOFSKY (1949) fitted a few cubic polynomials to the wind and contour fields over an area of the order of 10^6 square mi, which is about four times greater than the area used for Computation No. 1. Winds were weighted by a factor corresponding to $T = 2$. Visual comparison of the objective analyses with corresponding subjective analyses showed that the two methods gave similar contour patterns, but the objective analyses were somewhat smoother. Panofsky also computed the Laplacian derivatives of the subjective and objective contour fields. Here the agreement was "not good". The last result may have been anticipated, since the Laplacian of a cubic is a linear function and experience of vorticity fields computed from subjective analyses suggests that the true vorticity fields are very far from linear; in fact they are usually more complex than the contour patterns themselves. The implication of comparative simplicity of the vorticity field may be a considerable objection to the use of polynomials in representing contour patterns over large areas.

Various networks

Computation No. 4

The six-station network, centred upon Liverpool, was chosen somewhat arbitrarily as suitable for use in the above interpolation experiments. To examine the effect on the errors of representation of varying the distribution and quantity of data, quadratic surfaces were fitted to sixteen selections of data from the eight stations used in Computation No. 2. The computations were made with a weighting factor $T = 2^{\frac{1}{2}}$ for each of 40 occasions. The r.m.s. errors of representation at Liverpool and at stations used in the computations are given for each selection in Table 1, which also contains, for comparison, values previously obtained for corresponding occasions in the six-station and eight-station calculations (Computations Nos. 1 and 2).

The first six rows of Table 1 (series B, C, D, E, F and six-station) give information regarding the relative values of contour heights and winds. Comparing the Liverpool interpolation errors, series D, E and F, which contained more heights than winds, improved markedly upon series B and C in which winds preponderated.

Series M, K and I contained wind and contour height observations from five, four and three stations respectively, a compromise being made in the choice of stations between proximity to Liverpool and regularity of distribution. Each of these arrangements have errors of interpolation similar to those of the six-station data, except that the three-station arrangement (series I) gave a slightly better fit to the Liverpool wind: Liverpool is not far from the centroid of the triangle formed by three stations, Aldergrove, Hemsby and Camborne. When the size of the triangle was increased (series J and H), the interpolation errors were also increased.

Most of the above selections were chosen so as to have a reasonably symmetric distribution with respect to Liverpool, although comparison between series D and E shows how the disposition of the two winds materially affected the contour height errors. In series N, O and P the data were not symmetrically arranged about Liverpool, in fact for series O and P the problem was one of extrapolation rather than interpolation. In these last three cases the errors of representation were comparatively large.

Oceanic networks

The foregoing experiments gave promising results, but the British network of upper air observations is one of the most satisfactory for accuracy and density of observations. The worst networks are found over the oceans. A difficulty which arises in experimenting with a weather ship network is that the reservation of one observation for verification purposes alone seriously depletes the network. The alternative to checking the objective interpolations against observations is to compare them with values read from a conventionally analysed chart. This latter practice may be unsound however, particularly for experiments in which T is varied, since the conventional analysis is subjective and itself depends upon the analyst's opinion regarding the relative weights to be given to wind and height observations. To avoid these difficulties the five stations, weather ships "I", "J" and "K", Crawley and Stornoway were chosen as being equivalent to a mid-ocean network. Valentia observations were used to test the accuracy of the interpolated values.

Computation No. 5

Quadratic surfaces were fitted to observations from the five stations for 100 occasions chosen from the year 1954. Computation were made for T-values of 1, 2 and 4. The errors of representation appear in Table 2. The interpolation errors considerably exceeded those obtained in previous computations.

Computation No. 6

Computation No. 5 was repeated, replacing the quadratic surfaces by cubics. Table 2 contains the results. The Valentia errors are slightly smaller for the cubics than for the quadratics.

Assessment

The experiments described above show that in a close network of observations, a reasonably accurate objective interpolation for contour height and wind can be made. Comparisons with the errors of representation of the conventional contour analysis suggest that in regions of good data coverage, objective analyses will be at least as satisfactory as conventional analyses for numerical prediction. The ocean network computations suggest that only in limited regions will an analysis which takes no account of continuity in time and in the vertical, give very good results. There is no difficulty in principle in introducing an element of time continuity into the analysis for, as GILCHRIST and CRESSMANN (1949) point out, forecast values of height and wind can be taken into account in the least squares process on the same basis as observations. The weight given to a forecast value would naturally be much less than the weight given to an observation.[1]

The object of this series of computations was to derive interpolated heights and winds for a single grid point. Before extending the technique to get interpolations over a network of grid points one is led to consider whether it is better to represent the contour field over the whole network by a single function or whether a fresh function should be derived for each grid point. GILCHRIST and CRESSMANN

[1] A number of analyses have now been made by F. H. BUSHBY, for the North Atlantic using this technique. These will be published shortly.

(1949) remark that the latter alternative agrees with synoptic practice in considering data only from the immediate vicinity of the grid point in question. For the former alternative difficulties may arise due to the dependence of the expression for the contour height at a given grid point on data from distant regions. If a single function is used for representation over a wide area, the degree of complexity of the implied geostrophic vorticity field may be over-restricted, although this objection would be less valid for harmonic functions for example, than for polynomials. On balance it seems that the method which treats each grid point independently is likely to prove the more flexible approach.

Conclusions

1) In a close network of reliable observations, objective analyses are capable of representing the 500 mb winds with a standard error of about 10 kt. This figure compares favourably with the errors of representation, about 14 kt., of routine analyses for the same area.

2) In more open networks, such as occur over the Atlantic Ocean, provided that both winds and contour heights are available, winds can be interpolated with a standard error of about 16 or 17 kt. Errors of the conventional analysis are not known for such a region, but they are unlikely to be less than 16 or 17 kt, judging from the data available for the land area.

3) The optimum weight to be given to wind observations as compared with contour heights varies according to the nature of the interpolation. When interpolating for wind or contour height using quadratic surfaces the weighting factor is best chosen so that a discrepancy of 30 ft in contour height receives the same weight as a discrepancy of 10 to 15 kt in wind. Using cubic surfaces, a similar weight gives good results for contour height interpolations but the optimum weight for wind interpolation is obtained by equating a discrepancy of 30 ft in contour height to a discrepancy of 5 kt in wind.

4) The errors of interpolation are not very sensitive either to small variations in the relative weights given to contour heights and winds or to small variations in the number and disposition of the observations.

Acknowledgements

This paper is published by permission of the Director of the Meteorological Office. The least-squares solutions were calculated electronically using the Ferranti Mk. 1 machine at Manchester University. The computations were coded by F. H. Bushby and Vera M. Huckle. Benefit of discussion with J. S. Sawyer, G. A. Corby and F. H. Bushby is gratefully acknowledged.

REFERENCES

Bergthorsson, P., and Döös, B., 1955: Numerical weather-map analysis. *Tellus*, *7*, p. 329.

Gilchrist, B., and Cressman, G. P., 1954. An experiment in objective analysis. *Tellus*, *6*, p. 309.

Murray, R., 1954: On the accuracy of contour charts in forecasting upper winds. *Met. Office, London, Professional Notes*, *7*, No. 110.

Panofsky, H. A., 1949: Objective weather-map analysis. *J. of Meteorology*, *6*, p. 386.

On the Accuracy of 500 mb Analysis with Special Reference to Numerical Forecasting

By ROY BERGGREN, Swedish Meteorological and Hydrological Institute, Stockholm

(Manuscript received April 18, 1957)

Abstract

The result of an analysis experiment is discussed. From 29 different analyses of one and the same 500 mb chart certain parameters were calculated. The standard deviation of the height of the 500 mb surface had an average value of 15 gdm, determined over an area of 4,200 × 8,100 km. The correlation coefficient between the observed changes and the changes forecast with a numerical barotropic forecast varied between 0.53 and 0.77 (24h), 0.77 and 0.84 (48h), respectively. Some other verification parameters are given. Calculations of the geostrophic wind are carried through for all analyses, both at points, over areas and along certain tracks. The deviations due to the differences in analysis are sometimes considerable.

Introduction

During the last decade the problem of numerical methods in making upper air forecasts has received much attention. Quite a few papers have been published giving detailed information about numerical forecasts: CHARNEY, 1949; BOLIN, 1955; CHARNEY-PHILLIPS, 1953; SAWYER-BUSHBY, 1953; Staff Members, Department of Meteorology, University of Stockholm, 1954; BERGTHORSON ET AL., 1955; GATES ET AL., 1956; DÖÖS, 1956. In most of these papers the success or failure of a forecast is demonstrated by certain numerical values such as the correlation coefficient between the observed and forecast changes, the root-mean-square difference between the same quantities, the ratio between this root-mean-square value and the standard deviation of the observed changes and so on. Many authors have expressed their disbelief in these quantities and have asked for other methods of indicating the degree of success of an upper air forecast.

The numerical methods are tested in different ways and in all cases the need for a good measure of the success of the forecast is acute. In some places the methods are used on an operational basis because one wants to have an indication of how the methods work when faced with practical problems such as communication troubles, machine breakdown, personnel requirements and so on. Inherent in these tests are of course tests on the possibility of the numerical forecasts to indicate changes in the upper air pattern, breakdown of blocking highs, occurrence of cut-off lows etc. Sometimes the methods are used on specially selected cases, when one wants to know how an interesting change in the upper air pattern is handled by the atmospheric model used in the method in question. Such an interesting case was observed during the first part of November 1949, when a radical change of the circulation pattern over the North Atlantic and Europe occurred. This situation was first studied by the author for an entirely different reason (BERGGREN, 1952), but was later picked out as a good trial case. The author hopes to be able to give an account of this test in a later paper.

During the work with the test it became evident that some specific questions had to be discussed before the final work could be done. One of the problems can simply be stated as follows: What are the limits within which the so called known (observed) changes of the height of the 500 mb surface can vary? When calculating the correlation coefficient between observed and forecast changes one often treats the values of observed changes as fixed values that are known with all required accuracy and discusses then differences between observed and forecast changes that may be well within the limits of accuracy.

The question of how an upper air numerical forecast is influenced by the analysis used and of how much the forecast would be changed if another base analysis was used, has been discussed by some authors. NEWTON (1954) and BEST (1956) have shown examples of how radical a numerical forecast may be changed, if another base analysis is used.

During the work with the November 1949 case the author got quite convinced that different analyses could give significantly different verification figures (correlation coefficients, errors and so on), i.e. besides the above discussed effect of the base analysis on the forecast the effect of differences in analysis of the verification map has to be taken into account. For this reason an experiment was carried out along the following lines.

Method of analysis experiment

Out of some investigated days in the early part of November 1949 the seventh was chosen. For this day one 700 and one 300 mb forecast were available together with two 500 mb forecasts made according to slightly different schemes. For reasons that may be evident in a forthcoming paper about the result of the tests as mentioned in the introduction this was a great advantage. Furthermore, the situation over the Atlantic was rather clear and the resulting differences ought to represent minimum rather than maximum values. If a situation with a deep Atlantic low had been picked out and resulted in large difference the argument against the result might have been the following: a) the chosen situation is such that maximum differences could be expected and b) if another situation had been chosen, the differences might well have turned out to be insignificant. Finally, it was necessary for the further work to choose one of the days already studied.

The observations from the author's working chart were plotted on tracing paper and each station was checked thoroughly for plotting errors. Then ozalid copies were taken in order to give the analysts identical copies. In fig. 6 the stations used are indicated. The outer boundary was the limit of the analysis and stations were plotted a certain distance outside this boundary, which depended on the

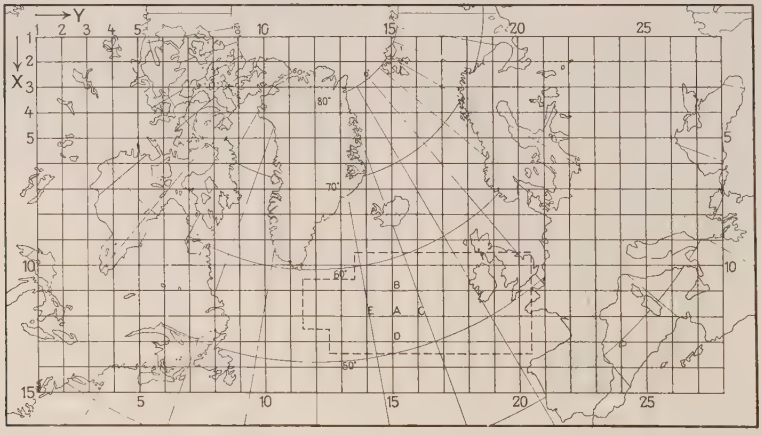


Fig. 1. Location of the grid used for computation of verification figures, geostrophic wind etc. The boundary indicates the region within which the analysis was performed.

amount of data available in the boundary area and varied between 1,000—3,000 km.

The analysts were contacted by the author one at a time and the chart to be analysed, the 500 mb for the 6th (fig. 3a) and the sea-level chart for the 7th (fig. 4) were given to them. The analyst was given instructions, containing among other things a statement that the analysis must not be discussed with other analysts before the trial period was over. In order to get as homogeneous a sample as possible all analysts were told to make "a careful routine analysis", i.e. not to spend too many hours on the analysis and use the same technique as they use in the daily analysis work. On an average the analysts worked with the analysis about two hours. One difficulty was that some of the analysts would have preferred to use the differential analysis technique, but this possibility was deliberately taken away by the author. The reason for this being that in a case like the one used with a rather simple analysis over the ocean the deviations between analyses ought to be smaller if all used a direct analysis method instead of having, say, 50 per cent of the analyses made in accordance with the differential technique. All analysts are accepted as analysts at the Swedish Met. and Hydr. Institute in Stockholm (SMHI).

The material on which the trial analysis was based consisted of reevaluated data. This means that transmission errors were eliminated. This type of error may reach any value and the possibility to trace the error is depending among other things on the density of the station network. The base material was better than the one used in the daily routine work and the expected deviations between the different analyses were thus reduced. In some cases there remained doubt as to whether a certain sounding was correct or not. This question will be discussed somewhat later on in the paper.

It was possible to get 26 different analyses within the time limit set for the trial period. When the analyses were handed over to the author, he personally interpolated height values in each of the 420 grid points indicated in fig. 1. When making such an interpolation it is inevitable that a personal error of interpolation comes in. This error was, however, minimized when all charts were read off by the same person. There was no indication that the interpolator learned some values by heart. On the contrary, it was astonishing how easy it was to avoid such an influence!

The author's working chart for the 7th was read off twice with nearly two years interval. The mean difference was 1.9 gdm[1], and the differences were distributed thus: zero difference in 206 points, ± 5 gdm in 164 points, ± 10 gdm in 49 points, ± 15 gdm in 1 point. The standard deviation was 4.3 gdm. It is very reasonable to assume that the standard deviation of interpolation varies with the isohypse gradient. The material at hand did not, however, permit an investigation of this question.

The first obvious thing to do with the collected analysis material was to calculate, at each grid point, the standard deviation of the height of the 500 mb surface (σ_h). Fig. 5 gives the result of these calculations together with the mean 500 mb chart, calculated from the 26 analyses. One thing is clear from a first sight on fig. 5 viz. the non-relationship between the standard deviation and the upper air pattern. The largest standard deviation, around 44 gdm, is observed over the North West Territories of Canada, near the radiosonde station 72938 Coppermine and at northeasternmost Greenland. Other maximum areas are found east of Newfoundland (38 gdm), at the southwestern shore of Hudson Bay (37 gdm), over the Black Sea (32 gdm), over Italy (29 gdm), over Baffin Island (28 gdm), south and southeast of the southern tip of Greenland (26 gdm), and over the outer Bay of Biscay (25 gdm). Minimum values are observed over northeasternmost USA and adjacent Canadian areas (2—3 gdm), the British Isles, eastern Russia and northeasternmost part of Baffin Island (3—4 gdm).

The overall mean of σ_h for all 420 grid points is 14.9 gdm. The values of σ_h ought to be corrected for the standard deviation due to the interpolation. This correction has not been applied in this case, however, because all charts discussed have been read off in the same way. Thus, all charts are affected in the same manner, if we disregard the variability due to different isohypse gradients. Furthermore, the technique of reading off values at grid points

[1] All through this paper the heights are given in geodynamic meters (gdm).

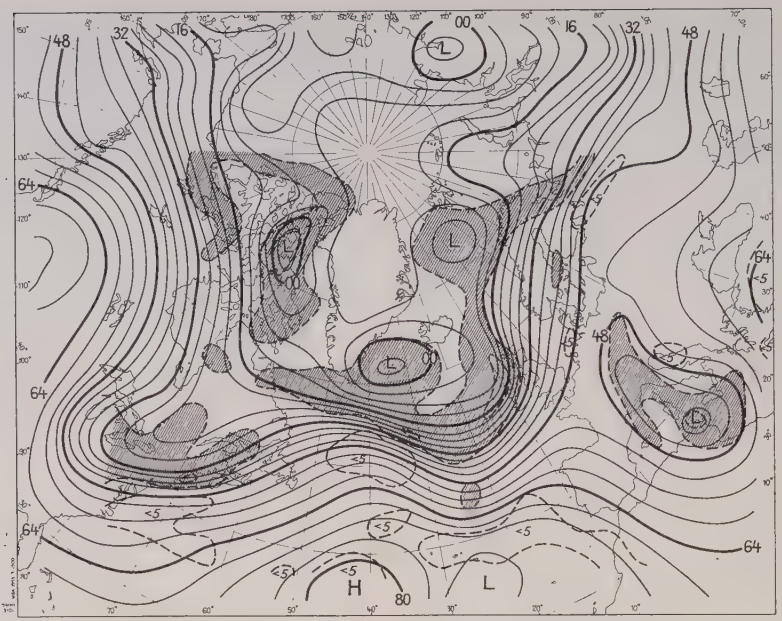


Fig. 2 a. 500 mb, 0300 GMT 5 November 1949, with isohypses for every 4 geodynamic decameters (solid lines) and isolines for the absolute vorticity in 10^{-5} sec^{-1} (dashed lines). Hatched areas indicate vorticity values larger than $15 \cdot 10^{-5}$ sec^{-1}. Isoline interval 10^{-4} sec^{-1}.

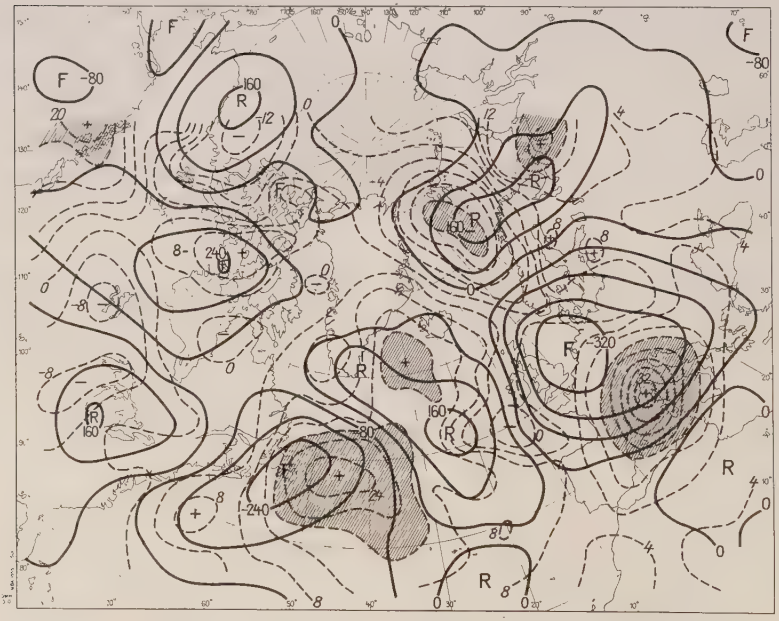


Fig. 2 b. The observed changes (solid lines; isallohypse interval 80 gdm) of the height of the 500 mb surface between 0300 GMT 5—0300 GMT 7 November 1949 and the difference between observed and forecast changes (dashed lines; isoline interval 4 decameters) for the same period. Hatched areas indicate values of the difference *numerically* larger than 16 geodynamic decameters. *F* and *R* indicate extreme values of the observed changes; plus and minus signs indicate extreme values of the difference between observed and forecast changes.

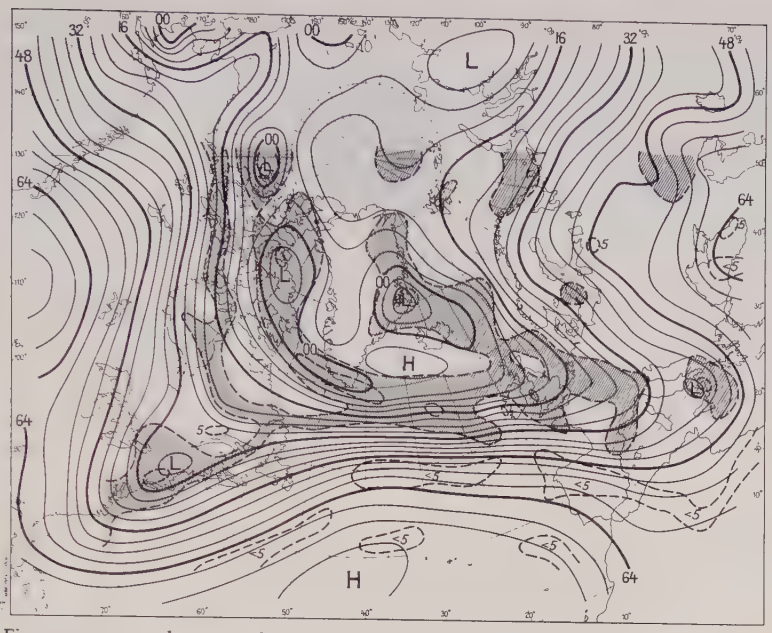


Fig. 3 a. 500 mb, 0300 GMT 6 November 1949, with isohypses for every 4 geodynamic decameters (solid lines) and isolines for the absolute vorticity in 10^{-5} sec^{-1} (dashed lines). Hatched areas indicate vorticity values larger than $15 \cdot 10^{-5}$ sec^{-1}. Isoline interval 10^{-4} sec^{-1}.

Fig. 3 b. The observed changes (solid lines; isallohypse interval 4 decameters) of the height of the 500 mb surface between 0300 GMT 6—0300 GMT 7 November 1949 and the difference between observed and forecast changes (dashed lines; isoline interval 4 decameters) for the same period. Hatched areas indicate values of differences *numerically* larger than 8 geodynamic decameters. *F* and *R* indicate extreme values of the observed changes; plus and minus signs indicate extreme values of the difference between observed and forecast changes.

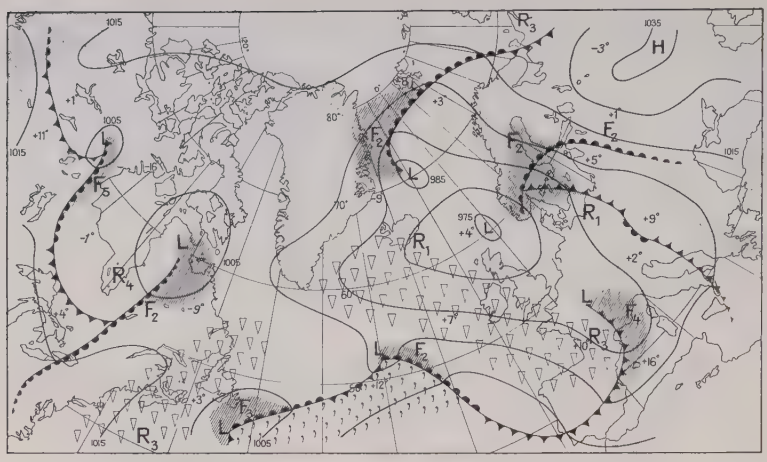


Fig. 4. Sea-level chart for 0000—0300 GMT 7 November 1949.

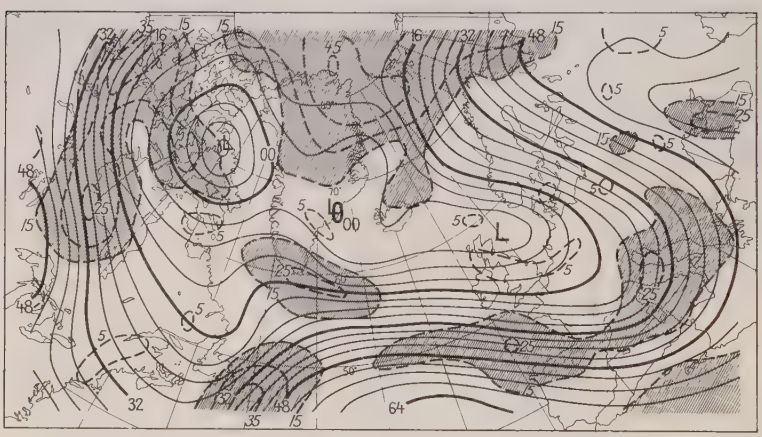


Fig. 5. Standard deviation (dashed lines) of the height of the 500 mb surface calculated from 26 different analyses (σ_h) based on identical material for 0300 GMT 7 November 1949. Hatched areas indicate values larger than 15 gdm. The average 500 mb chart calculated from the 26 analyses is given by the solid lines.

is the most common one so far. Numerical analysis technique eliminates this type of error.

The distribution of the standard deviation looks a little erratic and as stated above no relation with the isohypse pattern can be seen. It ought to be more fruitful to look for a relationship between the position and number of radiosonde stations and the standard deviation. Three different approaches were tried out. According to the first method the number of radiosonde stations in a square of 3×3 grid widths was counted and plotted at the central grid point, the grid width being 1.5 cm or approx. 300 km since the working chart was of the scale $1:2 \cdot 10^7$ (polar stereographic chart with the scale factor equal to one at 50° N). The maximum value of the station density, 14 stations per square, was observed over western Germany, the Netherlands and southeastern North Sea. The area with 6 stations or more per square covered Europe from the latitudes of the Alpes to northern Scandinavia and from the British Isles to eastern Russia. Within the grid area (fig. 1) the value of 6 or larger was also found over northeastern USA and at a few isolated points over the North West Territories of

Canada. A network density equal to zero was observed over northeasternmost part of Greenland, over Spitzbergen and the region between and north of those areas, over the central part of the Atlantic at the boundary of the grid, at one isolated point off the Newfoundland coast, at one point on the southwestern shore of Hudson Bay and finally at some grid points over Yugoslavia, Albania and Greece.

Comparing the distribution of the standard deviation as given in fig. 5 with the statements above about network density one can immediately see that the relationship between the standard deviation and the network density is high in some areas. At a first look one is therefore tempted to calculate a correlation coefficient between these variables. The coefficient turns out, however, to be as low as 0.31.

The explanation of the low coefficient is the fact that some areas have high values of network density together with high values of the standard deviation due to erroneous or questionable observations while for other areas the opposite is true due to the fact that the isohypse gradients are very weak in those areas. These two effects will be discussed later.

When plotting the pair of values at all grid points in a scatter diagram one can see that a straight line or any other curve is not able to take all pair of values reasonably into account.

Next, another network density was calculated taking upper wind stations into consideration. The simplest way possible to do this is to give radiosonde stations and wind stations equal weight. Other more elaborate schemes can be developed, but in this case they would lead nowhere. The isolines for the network density defined in the new manner coincide practically everywhere with the isolines discussed above. No improvement could be obtained this way.

If one tries to analyse how an analyst works when making an upper air analysis, one arrives to the conclusion that only the nearest-by stations are used when a certain area is analysed. Let us take one example: when the area just off the Scottish and Irish coasts is analysed the stations 03026 (Stornoway) on the Hebbrides and 03917 (Aldergrove) and 03953 (Valentia) on Ireland are used together with the two nearest weather ships but not at all or to a much lesser degree 03485 (Downham Market) and 03808 (Camborn) in England. The latter stations are mostly used as checks on the more near-by stations. This reasoning leads one to the conclusion that the important factor is not the network density but rather the distance from the analysed area to the nearest station. When applying this reasoning to the problem at hand, one encounters the difficulty that has already been discussed. Which is the best way to take both radiosonde stations and wind stations into account? In order to get a first idea of the important factors the wind stations were disregarded. This is not as bad as it may look at a first glance. Over large areas where the number of radiosonde stations is low as over the Atlantic, a radiosonde station quite often is a wind station at the same time. On the other hand a single wind station in a vast area without radiosonde stations is not as valuable as an additional radiosonde station.

The simple scheme used was to determine the distance from each grid point to the nearest radiosonde station. The distribution of this variable (R) is given in fig. 6.

The overall maximum value (50 mm $\simeq$ 1,000 km) is found at the northeasternmost tip of Greenland and in the lower right corner of the grid. The highest value over the Atlantic (42 mm $\simeq$ 840 km) is obtained in the bottom row and in the column that passes through Spitzbergen and Iceland. Over the Adriatic Sea R reaches the value 35 mm ($\simeq$ 700 km) and just southwest of Hudson Bay 32 mm ($\simeq$ 640 km). The maximum value east of Newfoundland is 37 mm ($\simeq$ 740 km) and southeast of Greenland 30 mm ($\simeq$ 600 km). If one compares the distribution of R with the standard deviation (σ_h) for the 26 different analyses, one can immediately see, that a rather high correlation is at hand. A calculation of the correlation coefficient between these two variables gives the result 0.90. The equation for the regression of the standard deviation (σ_h in gdm) on the distance from a grid point to the nearest radiosonde station (R in mm; 1 mm corresponding to approx. 20 km) reads as follows

$$\sigma_h = 0.84\ R + 3.2$$

The standard error of estimate is 3.1 gdm. It is of interest to investigate where the above

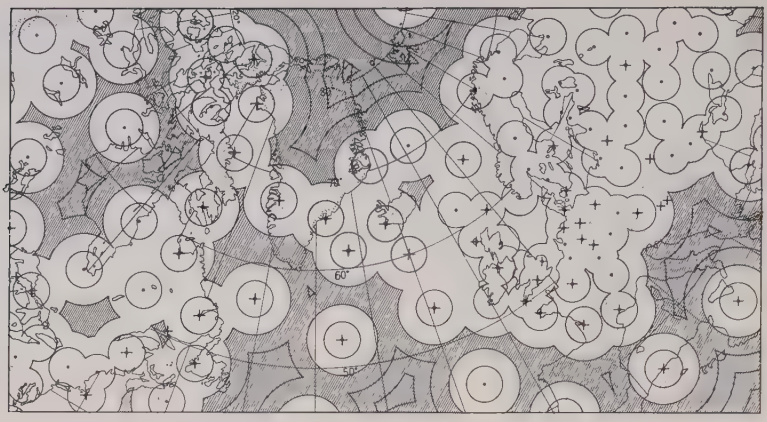


Fig. 6. The distribution of the parameter R, defined as the distance (in mm, 1 mm $\simeq$ 20 km) on the working chart, a polar stereographic chart (1 : 20 mill.), from a certain point to the nearest-by radiosonde station. The circles are drawn for R equal to 10 mm, 20 mm, 30 mm and so on. Hatched areas indicate values of R greater than 20 mm ($\simeq$ 400 km). Dots and crosses give the positions of radiosonde stations and wind stations, respectively.

Fig. 7. The distribution of the difference between observed (σ) and estimated values of the standard deviation of the height of the 500 mb surface. Hatched areas with lines sloping from lower left to upper right corner indicates differences numerically larger than 10 gdm. Hatched areas with lines sloping from lower right to upper left corner indicate values of σ_h larger than 15 gdm (cf. fig. 5). Plus and minus signs indicate extreme values of the difference.

regression equation is valid and in which areas it gives large errors. In fig. 7 the difference between the observed and the estimated standard deviation is given. Except for some small areas containing only one grid point an absolute value of this difference greater than 10 gdm is observed over the following areas: the North West Territories of Canada (maximum value +37 gdm), the Canadian archipelago (+15 gdm), southernmost Baffin Island and areas south thereof (−12 gdm), the central Atlantic (−26 gdm), Spain (−15 gdm), Tunisia (−22 gdm), Greece and Yugoslavia (−15 gdm), the Black Sea (−24 gdm) and finally northern Finland and the Kola Peninsula (+12 gdm).

Large positive differences occur where large values of the standard deviation are found near radiosonde stations. One marked area of this type is situated over the North West Territories of Canada (figs. 6 and 7). Here a grid point as near as 4 mm ($\simeq$ 80 km) from a

radiosonde station has a standard deviation of about 44 gdm.

The reason for this peculiar discrepancy can be found in the fact that the radiosonde station in question is one of the very few on the actual 500 mb chart, the height value of which may be questioned. About fifty per cent of the analysts considered the value to be too low and made a smoothed analysis in this area. A thorough check of this station has shown that the height value may be taken as correct.

The large difference observed over the Black Sea is due to another type of difficulty. The Crimean station had a 500 mb value that definitely was too high. Some of the analysts disregarded the station altogether while others tried to take it into account as much as possible with a varying degree of smoothing.

The small positive area over the Canadian archipelago may be explained by the fact that this area is situated at the outskirts of vast areas around the north pole completely lacking stations. Even if the grid points in question are situated not too far away from a radiosonde station the lack of stations north and west of the area puts the analyst into difficulties.

The area with positive differences between observed and calculated standard deviations that is found over northern Finland and the Kola Peninsula is due to the fact that the height values given by the stations 02836 (Sodankylä) and 22113 (Murmansk) were incompatible. Some of the analysts tried to take both stations more or less into account while others struck out one of them, the result being an increased value of σ_h.

The four areas with negative differences between the observed and estimated standard deviations in the lower right corner of fig. 7 may all be explained in the same way. The upper air pattern in these regions is rather clear cut and in some cases the gradients are not too strong. It is not too difficult to get a good picture of the isohypse configuration near the radiosonde stations in the lower right corner. It is then easy to extrapolate this pattern into areas with scattered data.

We have now discussed the areas with large differences between observed and calculated values of the standard deviation. It is, however, interesting to note in fig. 7 that practically all areas with a standard deviation equal to or larger than 15 gdm fall outside the areas discussed above. As a whole one can state that the estimated values of the standard deviation coincide rather well with the observed ones, the mean deviation (i.e. the average value of the absolute values) being 5.4 gdm.

A comparison between fig. 6 and 7 shows that the absolute value of the difference between observed and estimated standard deviation generally does not exceed 10 gdm in areas with a station distribution such that no grid point is more than 20 mm ($\simeq$ 400 km) away from a radiosonde station.

The discussion so far has given some figures of standard deviations, that are due to different analyses of one and the same observation material. These values may be compared with the standard error of the radiosoundings. According to NYBERG (1952) the standard deviation of height values (500 mb) between two radiosondes was about 30 gdm, a value found during the international radiosonde comparisons at Payerne in 1950. Before this standard deviation was calculated the differences between the radiosonde values had been corrected for systematic differences. The average standard deviation of height values for 500 mb given by a single sonde was about 18 gdm. It is very difficult to apply these values in one case because in some areas, such as the British Isles and the United States of America, the radiosonde network is rather homogeneous while in others, such as Central Europe, the network is composed of a great variety of radiosonde types. A reasonable overall mean value of the standard deviation of the height of the 500 mb surface originating from instrument errors of different types might be 25 gdm. The same numerical value of the standard deviation for the 26 analyses corresponds to a distance of 26 mm between a grid point and the nearest-by radiosonde station (R) i.e. a distance of about 520 km at the surface of the earth. On the other hand the observed mean value of R corresponds to a standard deviation of 17 gdm. Thus, the deviations due to different analyses attain values that cannot be neglected in comparison with the instrument errors.

Finally it might be of interest to investigate the distribution of the range of the height values (i.e. the difference between the highest

Table 1. Frequency distribution of the range of the height of 500 mb surface originating from 26 different analyses of the upper air data for 0300 GMT 7 November 1949. The range is determined at each of the 420 grid points in Fig. 1

range	number of points
0—20 gdm	27
21—40	149
41—60	93
61—80	54
81—100	46
101—120	25
121—140	14
141—160	7
161—180	2
181—200	3
Total	420

and the lowest value) as interpolated from the 26 analyses at each grid point. This distribution is given in table I. The range varied between 10 and 195 gdm. The maximum value was observed at the north-easternmost tip of Greenland and the minimum value over the Irish Sea and at the grid point over the northernmost part of Baffin Island ($x=8$, $y=4$). The mean value was 60 gdm.

It ought to be emphasized here that the investigated analyses were based on checked data without transmission errors. Furthermore, it has been stated earlier that the upper air pattern was not a difficult one to analyse.

Verification of numerical forecasts

The main object of the analysis experiment discussed in this paper was to see how among other things verification figures for numerical forecasts were changed if another verification analysis was used. In order to arrive at this goal a number of verification figures were computed using a high speed computer (BESK in Stockholm).

The following variables were computed:

$\bar{X}$ the mean value of the 24h and 48h observed changes (x) of the height of the 500 mb surface from the 5th and 6th, respectively, to the 7th of November 1949

$\overline{|X|}$ the mean deviation of x

σ_x the standard deviation of x

ε_0 the root-mean-square of the difference between observed and forecast changes

r_{xy} the correlation coefficient between observed and forecast changes

$\overline{d_0}$ the mean value of the geostrophic wind that corresponds to the observed changes

$\overline{d_e}$ the mean value of the geostrophic wind that corresponds to the error; i.e. the difference between verification analysis and forecast

δ_0 the root-mean-square of d_0

δ_e the root-mean-square of d_e

$(d_0)_{\max}$ the maximum value of all the d_0's

$(d_e)_{\max}$ the maximum value of all the d_e's

$r_{d_0 d_e}$ the correlation coefficient between geostrophic wind values corresponding to the observed and forecast changes.

All the above values were calculated for the 26 analyses over the grid in fig. 1, which contains 420 points. The maximum and minimum values for each variable were picked out. These extracted values are given in table II. In addition to the 26 analyses the author had access to a routine analysis made at the Swedish meteorological and hydrological institute, his own analysis, and a numerical analysis made according to the scheme described by Bergthorson and Döös (1955) and Döös and Eaton (1957).

The first two analyses were made with differential analysis technique, more completely carried through in the latter case.

Verification figures were calculated for two different numerical barotropic forecasts, one 24h forecast and one 48h forecast, both forecast periods ending at 0300 GMT on the 7th of November. Thus, the two forecasts were not based on the same analysis. The forecasts were made according to the methods described by Staff Members, Dept of Met., Univ. of Stockholm (1954) and Bolin (1955).

Figs. 2a and 3a give the 500 mb chart for the 5th and the 6th, respectively. The observed changes of the height of the 500 mb surface are given in fig. 2b and 3b. In the two latter figures the errors, i.e. the difference between observed and forecast changes also are given.

The values of some observed quantities as given in table II have to be compared with the corresponding forecast values. These are given in table III.

a) 24h forecast

Already in the mean value of the observed changes we observe a rather large variability.

Table II. Some verification parameters computed from 29 different verification 500 mb analyses (0300 GMT 7 November 1949), a given 500 mb chart for 0300 GMT 6 November 1949 and a numerical barotropic forecast. The lower part gives the corresponding values for the 48h forecast based on the 500 mb chart for 0300 GMT 5 November 1949. Values of maximum, minimum, mean and range are taken from 26 analyses

24h	$\overline{X}$	$\overline{\lvert X \rvert}$	σ_x	ε_0	r_{xy}	δ_0	δ_e	$\overline{d_0}$	$\overline{d_e}$	$(d_0)_{\max}$	$(d_e)_{\max}$	$r_{d_0,\, d_p}$
max.....	—26	65	76	68	0.651	12.0	11.0	10.6	9.6	35.2	30.6	0.485
min.....	—38	55	63	57	0.564	10.8	9.4	9.7	8.3	25.3	23.6	0.343
mean.....	—32.1	60.6	70.0	63.7	0.609	11.5	10.2	10.2	9.0	29.1	27.5	0.414
range.....	12	10	13	11	0.087	1.3	1.6	0.9	1.3	9.9	7.0	0.142
routine...	—27	58	66	60	0.609	11.6	10.5	9.9	8.7	33.6	38.2	0.465
author....	—32	62	71	69	0.534	12.4	11.4	10.6	9.7	36.5	33.6	0.432
NA.....	—27	55	63	45	0.772	10.2	6.9	8.8	6.0	30.7	15.8	0.698
48h												
max.....	—59	114	127	98	0.838	18.8	12.4	16.7	10.7	46.8	36.6	0.731
min.....	—72	104	118	88	0.772	17.8	10.9	15.5	9.5	42.8	25.1	0.630
mean.....	—65.5	109.2	124.0	94.1	0.810	18.5	11.6	16.1	10.1	44.5	28.4	0.676
range.....	13	10	9	10	0.066	1.0	1.5	1.2	1.2	4.0	11.5	0.101
routine....	—61	106	121	90	0.812	18.7	12.8	15.7	10.4	55.4	45.5	0.626
author....	—65	109	125	98	0.783	19.0	13.0	16.3	11.0	49.7	34.1	0.622
NA.....	—60	102	118	90	0.804	17.5	10.6	15.2	9.1	44.1	28.1	0.718

The deviation from the overall mean value (-32.1 gdm) is ± 6 gdm or 19 per cent. The deviation of an individual value from the mean of the standard deviation may be 10 per cent. The root-mean-square of the error may vary within a range that is 17 per cent of the mean value.

Quite often one is not interested in the correlation coefficient, but rather in its squared value. The extreme values for this quantity is 0.318 and 0.424, respectively, a rather large variation.

From table II we can also see that the geostrophic wind field corresponding to the error has a mean value that is nearly as large as the one corresponding to the observed changes. Furthermore, the range is 14 per cent of the mean value. The maximum value at any grid point of these two quantities varies considerably, especially the maximum value of the wind corresponding to the observed changes.

As already mentioned above the author has had access to three additional analyses. The first one is the routine analysis made at SMHI on 7th November 1949. The second analysis was made by the author in 1951. It is interesting to notice that these two analyses that were made with the differential analysis technique differ considerably in some respect from the 26 analyses discussed so far. From table II we find that the author's analysis gave a correlation coefficient of 0.53 which ought to be compared with the maximum value 0.65. Using Fischer's z-transformation we find that the probability to get such a difference out of chance is 6 in 100.

The low correlation coefficient obtained with the author's analysis is explained in the following way. The trough over the British Isles, Central Europe and Italy on the 6th (fig. 3a) moved very little during the next 24 hours (cf. fig. 5), but deepened somewhat. The forecast, however, brought the southern part of the trough to the northeast, thus causing a forecast error in the Mediterranean area. The author's analysis has the lowest value in the bottom of the trough of all analyses, where, by the way, eleven of the twentysix analyses showed a closed center. The difference between the verification chart and forecast chart obtained in this area using the author's analysis is larger than all those obtained with the twentysix analyses. As the changes in this area play an important role when one calculates the correlation coefficient for the whole verification area, this results in a low value of the correlation coefficient when the author's analysis is used.

Table III. Some forecast values for the 24h (48h) forecast based on the 500 mb of the 6th (5th) of November 1949, corresponding to the observed values given in table II

	24h	48h
$\bar{y}$	— 12 gdm	— 10 gdm
$\overline{\lvert y \rvert}$	51 »	97 »
σ_y	64 »	122 »
$\bar{d}_p$	9.4 m/s	15.7 m/s
δ_p	11.0 »	18.1 »

Real large differences were introduced in table II, when the values obtained with the aid of the numerical analysis were entered. This analysis is obtained by using among other approximations the 24h forecast ending at the time of verification (see BERGTHORSON and DÖÖS, 1955). When verifying the forecasts with a numerical analysis constructed in the manner developed by the above-mentioned authors, one runs the risk of being a victim of an "a(a+b)"-effect, i.e. one correlates two variables, the one of which to some extent is contained in the other [mentioned by DÖÖS (1956)]. In the special case discussed in this paper, this effect is obvious. While the correlation coefficient for the 26 analyses varies between 0.56 and 0.65, the numerical analysis gives the value 0.77. The mean value of the geostrophic wind corresponding to the difference between forecast and verification analysis ranges from 8.3 to 9.6 m/s, while the numerical analysis gives the very low value 6.0 m/s. The maximum value of this variable varies between 30.6 and 23.6 m/s for the 26 analyses while it is as low as 15.8 m/s for the numerical analysis.

It is interesting to notice that it was possible to increase the value of the correlation coefficient from 0.53 to 0.65 by choosing another analysis and as high as to 0.77 if the numerical analysis is taken into account. The square of these values 0.285, 0.424 and 0.596, respectively, makes the differences look even more discouraging.

The fact is that discussions on the reliability of the numerical forecasts are sometimes based on correlation coefficients and suchlike calculated from conventional verification analyses and sometimes on the same variables calculated using numerical analysis. One has to be very careful while doing so.

The variability of the correlation coefficient between observed and forecast height changes of the 500 mb surface is to some extent depending on the absolute value of the changes. When the changes are small the correlation coefficient may react sluggishly to the variation of the fields of observed and forecast changes originating from different analyses of the same observation material. Thus, it is of interest to investigate, whether or not the situation studied in this paper represents a case with extreme values of observed changes.

The Staff Members, Inst. of Met., Univ. of Stockholm (1954), BOLIN (1955) and DÖÖS (1956) have published figures showing that the mean value given in this paper is not a markedly low one. The values are in most cases for the same season.

b) 48h forecast

The height changes for the 48h period ending at 0300 GMT on the 7th of November were certainly larger than the 24h ones. In the lower part of table II the verification figures for this period are given and in table III the corresponding forecast values are found.

In table II we notice that the correlation coefficients are markedly higher for the 48h forecast than for the 24h forecast. The mean value of the standard deviation of the observed changes was about the same as the ones given by BOLIN (1955) and DÖÖS (1956). The ranges for X, $\overline{|X|}$ σ_x and ε_0 are of about the same value as for the 24h forecast. As the mean values, however, are higher for the 48h forecast, the percentage errors are lower. The range of the correlation coefficient between observed and forecast changes are somewhat lower in this case.

The probability to get a difference between correlation coefficients as the one observed here (0.77 and 0.84, respectively) is about 5 in 100.

For the different wind fields the absolute values of the ranges are about the same as for the 24h forecast and in some columns they are also the same relatively speaking.

There are remarkable differences in some columns between the extreme values obtained from the 26 analyses and the ones obtained from the routine analysis and the author's analysis. This is especially true for the maximum value of d_0 and d_e. The numerical anal-

ysis does not, however, differ as markedly from the others, as was the case with the 24h forecast. In this case there is no "a(a+b)"-effect, since the 48h forecast is not directly contained in the numerical analysis. In many columns the numerical analysis shows values within the range of the 26 analyses and the deviations are mostly small.

The conclusion is that for some verification figures in table II the 48h forecast gave lower relative values than did the 24h forecast. In other cases, however, the relative values are about the same and this is also true for the absolute values of the ranges in many columns. It would have been very interesting to widen this investigation to include many more cases. This is, however, impossible for the author because of the great amount of work involved in each case. It has to be emphasized once again that the results obtained in this investigation are only directly valid for these special cases and cannot be used for other cases without much care and further investigation. The purpose of this paper is only to give examples of what may happen due to variability in the analysis of one and the same observation material, i.e. the differences due to the personal factor in chart analysis.

It also has to be emphasized that the sometimes large differences obtained in this investigation are due only to the variability in the analysis of the chart at one end of the period. In the general case both analyses on which the calculation of the observed changes and so on are based are affected. Take for instance one grid point over the Atlantic somewhere halfway between two weather ships. A reasonable value of the standard deviation of the height of the 500 mb surface due to analysis effect may be 25 gdm. Let us further assume that the standard deviation on the two charts are independent of each other. This assumption is a reasonable one, because the analysis differences at one grid point as obtained by one analyst only to some degree may be considered systematic. Furthermore, in the daily routine work the analyses of two consecutive 500 mb charts, 24h apart, are often analysed by different persons. Under these assumptions the standard deviation of the observed changes is about 35 gdm. A standard deviation of the height of 30 gdm gives the value 42 gdm.

Differences in the total wind component along certain tracks

The discussion so far in this paper has been confined to the absolute height values of the 500 mb surface and to the 24h and 48h changes, observed and forecast. Taking the importance of air navigation into account as well as the necessity in many forecast techniques of knowing the upper winds accurately, it would be of interest to investigate the variability of the geostrophic wind due to different analyses.

For that purpose the geostrophic wind field was calculated for each analysis and for each grid point. The calculation was made at the same time when the above-mentioned verification figures were calculated. At each grid point the x- and y-components (see fig. 1) of the geostrophic wind were obtained. The components were computed over two grid widths, i.e. over a distance of about 600 km. This had the effect that no wind values could be obtained for the points along the boundary of the grid. Furthermore, it was technically an advantage to strike out the bottom row. The result was that the wind field was obtained over a grid consisting of 26×12 points.

In order to get a view of the influence of the differences in the analyses on the geostrophic wind field a number of summations were made. The wind components in the x-direction (v_x) were summed in the x- and y-direction (v_{xx} and v_{xy}), respectively. The y-components were treated accordingly. All these sums for the twentysix analyses were put together in four tables and among these values maximum and minimum values were picked out for each column (x) and row (y). These extreme values were divided by the number of points in the row and column, respectively. The resulting extreme values of the mean geostrophic wind components along rows and columns were entered in table IV and V. The mean values and the ranges for the different rows and columns are also given. In the same tables are also included the corresponding values for the routine analysis, author's analysis and the numerical analysis.

The largest column value of the range of the mean x-component of the geostrophic wind (v_{xx}) is 8.9 m/s[1] and for the y-component

Table IV. Geostrophic wind components along the x- and y-axes (v_x and v_y, respectively, see fig. 1[1] range are taken from 26

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\overline{v_{xx}}$	max.....	24.8	29.2	23.7	18.1	11.4	9.0	4.0	—4.9	— 5.6	—0.5	3.8	2.1	—1.2
	min.....	19.6	20.3	17.4	10.7	7.7	4.3	—0.9	—8.9	—11.0	—6.6	—1.5	—2.6	—5.4
	mean....	22.1	24.7	21.2	14.3	9.4	6.6	1.0	—7.0	— 8.1	—3.3	0.6	—0.2	—3.4
	range....	5.2	8.9	6.3	7.4	3.7	4.7	4.9	4.0	5.4	6.1	5.3	4.7	4.2
	routine ..	21.3	22.2	20.9	16.1	10.4	6.1	0.5	—6.0	— 6.9	—3.4	—1.2	—0.2	—2.5
	author...	23.4	22.7	19.1	15.5	11.7	7.5	1.4	—7.6	—10.3	—4.2	0.2	—0.3	—2.5
	NA.....	20.8	22.8	21.6	17.1	12.3	6.7	0.2	—5.8	— 6.5	—2.3	—0.8	—1.1	—3.3
$\overline{v_{yx}}$	max.....	—0.1	0.5	2.1	3.4	3.5	3.1	3.7	5.2	5.9	6.9	8.9	10.3	10.7
	min.....	—0.9	—1.0	—0.5	0.5	1.0	1.9	2.7	3.0	3.6	4.7	6.3	7.2	7.8
	mean....	—0.5	—0.2	1.1	2.4	2.7	2.8	3.3	4.0	4.7	6.1	7.5	8.6	9.0
	range....	0.8	1.5	2.6	2.9	2.5	1.2	1.0	2.2	2.3	2.2	2.6	3.1	2.9
	routine ..	—0.4	0.2	1.2	2.4	3.0	3.3	4.1	4.7	5.6	6.7	7.7	8.8	9.4
	author...	—0.3	—0.2	0.3	0.9	1.8	2.5	3.5	3.8	4.3	5.4	6.9	8.2	8.7
	NA.....	.0	0.3	1.2	2.6	2.9	3.4	3.8	4.8	5.3	6.8	7.5	7.9	7.9

summed in the same direction (v_{yx}) 3.7 m/s. The largest value of the range of the values of the y-component averaged along a row (v_{yy}) is 6.3 m/s and the one for the corresponding crosscomponent (v_{xy}) 1.4 m/s. It ought to be emphasized that the columns represent a distance on the surface of the earth of approximately 3,300 km, while the rows correspond to 7,500 km. The standard deviation of the mean geostrophic wind for a single analysis is 1.0 m/s for the column containing the maximum range of the x-component. It may be of interest to translate the above mean values into variations in flight time for an aircraft flying at a certain constant speed relative to the air along the tracks given by the rows and columns in fig. 1.

The concept of equivalent head wind may be used in this connection (SAWYER, 1950). In the formula for this theoretical wind the wind component across the flight path may be neglected for moderate values of that component. With a reasonable accuracy the wind retarding effect of a head wind may be taken in tables IV and V. In table IV, column 3, the maximum difference is 8.9 m/s. With a cruising speed of 100 m/s of the aircraft the delay (or gain of time) is 8.9 %. The flying time in a calm atmosphere is about 9.2 hours for a distance corresponding to a column. Thus, the delay is about 48 minutes, corresponding to a considerable operating cost of a modern large aircraft. For a row the corresponding values are: 6.3 m/s (row 12), 20.8h and 1.3h.

It is very reasonable to assume that the error of upper wind forecast is at least as large as the differences that can be obtained by using different analyses. Taking the case investigated in this paper into account and furthermore recognizing the fact that this case represents a rather simple analysis, it seems unlikely to arrive at an error of the 500 wind forecast that is less than 10 per cent on an average.

In tables IV and V it is interesting to notice that the numerical analysis in most columns and rows gives mean winds within the limits given by the 26 analyses.

The estimation of the geostrophic wind at certain points

Sometimes it is necessary to know the accuracy of the geostrophic wind at a particular point. In order to get an estimate of the magnitude of the standard deviation for the situation discussed in this paper, fourteen points were selected and the standard deviation (σ_g) of the absolute value of the geostrophic wind was calculated at each point. The coordinates of the points are given in table VI. Seven of the points represent grid points

[1] The wind values given in this paper are consistently 2 per cent off compared to correct values, the reason being that geodynamic meters were used instead of geopotential meters.

1) averaged along the x-axis (v_{xx} and v_{yx}, respectively). Values of maximum, minimum, mean and analyses. Wind speed in m/s.

15	16	17	18	19	20	21	22	23	24	25	26	27
—3.2	—5.3	—4.7	—1.9	—1.2	—4.6	—6.2	—6.3	— 6.6	— 8.3	—10.8	—11.3	—12.6
—8.4	—8.5	—7.4	—4.9	—4.1	—6.8	—8.3	—8.8	—10.6	—14.8	—14.7	—15.8	—17.7
—5.5	—6.8	—5.9	—3.3	—2.8	—5.5	—7.4	—7.5	— 8.7	—11.3	—13.0	—13.9	—14.6
5.2	3.2	2.7	3.0	2.9	2.2	2.1	2.5	4.0	6.5	3.9	4.5	5.1
—6.1	—7.4	—4.7	—2.9	—4.3	—7.1	—7.3	—5.4	— 5.9	—10.7	—16.9	—21.3	—23.4
—4.1	—5.8	—5.9	—3.6	—3.3	—6.7	—8.7	—6.3	— 3.3	— 5.9	—14.4	—20.0	—19.5
—5.6	—6.8	—6.0	—3.9	—4.0	—6.1	—8.6	—8.6	— 8.6	—12.1	—15.9	—15.4	—12.6
10.1	8.9	7.3	5.7	4.8	3.2	1.7	1.0	0.4	— 0.1	— 0.2	1.3	2.8
7.4	6.3	5.1	3.5	1.1	—0.1	—0.7	—0.8	— 1.1	— 1.5	— 1.4	— 0.4	— 0.3
8.8	7.8	6.3	4.6	2.9	1.6	0.7	0.3	— 0.3	— 0.8	— 0.6	0.2	0.8
2.7	2.6	2.2	2.2	3.7	3.3	2.4	1.8	1.5	1.4	1.2	1.7	3.1
9.4	8.5	7.3	6.1	5.3	4.0	2.5	0.5	— 1.4	— 2.3	— 1.6	0.1	2.2
8.6	7.8	6.7	5.0	3.6	2.1	0.8	0.1	— 0.6	— 1.2	— 1.4	— 1.0	— 0.7
7.7	6.8	5.2	3.9	2.9	1.9	0.9	0.2	— 0.2	— 0.7	— 0.3	0	0.1

with maximum value of the standard deviation of the height of the 500 mb surface (σ_h) in the corresponding areas (the standard deviations are given in table VI). The remaining seven points correspond to minimum values of the same standard deviation (cf. fig. 5).

The overall mean value of σ_g was 2.3 m/s. The mean value of the geostrophic wind itself was 21.2 m/s. Notice that the points ($x=2$, $y=5$; $x=4$, $y=27$) have large values of σ_h in spite of the fact that R is small. These two points have already been discussed in this paper and the discrepancy explained. The reason was that some of the analysts rejected the near-by radiosonde stations while others did not. This fact must have a repercussion also on σ_g and in table VI we can see that one of the points in question has the highest value among all the points.

From table VI we can see that there is a relationship between σ_g and the distance from the grid point to the nearest radiosonde station (R). The material is much too small to permit a calculation of a partial regression equation or suchlike. Theoretically, it ought to be a positive correlation between σ_g on the one hand and R and the absolute value of the geostrophic wind on the other. The more scarce the station network and the tighter the isohypse gradient, the more variable is the analysis. In table VI these effects may explain why the point ($x=2$, $y=13$), with $R=46$ mm ($\simeq$ 920 km), has the lowest value of σ_g among the points with large values of R. The gradients in this area were weak. Furthermore, the point ($x=13$, $y=13$) has a σ_g-value that is 1 unit larger than the corresponding value for the grid point ($x=13$, $y=5$). Its geostrophic wind speed is twice that of the latter while R is the same.

It is interesting to note that in two cases the range is as large as 16 m/s. The overall mean value is 9 m/s.

Finally an area over the Atlantic was selected as it represented a region with a broad, rather homogeneous air current. The area contained 33 grid points and is indicated in fig. 1. The grid points were divided into three groups according to their value of R (cf. fig. 6): $R \leq 10$ mm ($\simeq$ 200 km) (10 points), $10 < R \leq 20$ mm ($\simeq$ 200—400 km) (13 points), and $R > 20$ mm ($\simeq$ 400 km) (10 points). For each group the mean geostrophic wind speed was calculated from the 26 analyses together with the standard deviation of that mean value. As the selected area showed a current quasi-parallel to the y-direction of the grid (see figs. 1 and 5), the calculation was restricted to the wind component in that direction. The mean value of the cross-component was -1.0, 0.5 and -0.1 m/s, respectively, showing that this component could safely be disregarded, since the mean y-component was 28.0, 33.9 and 30.6 m/s, respectively. The standard deviation for the three groups was

Table V. Geostrophic wind components along the x- and y-axes (v_x and v_y, respectively, see fig. 1) averaged along the y-axis (v_{xy} and v_{yy}, respectively). Values of maximum, minimum, mean and range are taken from 26 analyses. Wind speed in m/s

	Row	1	2	3	4	5	6	7	8	9	10	11	12	13
v_{xy}		max. . . .	—0.8	—1.2	— 1.1	— 1.2	—0.8	—0.5	0.1	0.4	0.2	— 0.5	— 1.0	— 1.3
		min.	—1.5	—1.9	— 2.5	— 1.8	—1.4	—1.3	—1.2	—0.9	— 0.7	— 0.9	— 1.7	— 2.6
		mean . . .	—1.1	—1.5	— 1.7	— 1.4	—1.1	—0.9	—0.7	—0.4	— 0.4	— 0.8	— 1.4	— 1.8
		range . . .	0.7	0.7	1.4	0.6	0.6	0.8	1.3	1.3	0.9	0.4	0.7	1.3
		routine .	—1.3	—1.6	— 2.0	— 2.1	—1.9	—1.8	—1.6	—1.3	— 1.2	— 1.2	— 1.9	— 2.9
		author . .	—1.4	—1.6	— 1.7	— 1.5	—1.3	—1.1	—0.9	—0.6	— 0.5	— 0.9	— 1.4	— 1.5
		NA	—1.4	—1.6	— 1.6	— 1.5	—1.2	—0.9	—0.8	—0.7	— 0.8	— 1.0	— 1.4	— 2.0
v_{yy}		max. . . .	—0.7	—3.8	— 6.8	— 6.8	—6.1	—4.0	0.1	6.0	10.3	18.0	26.5	27.9
		min.	—4.4	—8.0	—10.6	—11.4	—9.2	—6.7	—2.9	1.6	6.1	13.8	20.2	23.3
		mean . . .	—2.6	—5.6	— 8.8	—11.3	—7.5	—5.8	—1.4	4.7	9.9	16.2	22.8	25.7
		range . . .	3.7	4.2	3.8	4.6	3.1	2.7	3.0	4.4	4.2	4.2	6.3	4.6
		routine .	—0.9	—2.7	— 6.3	— 9.1	—9.3	—7.2	—2.8	0.1	9.4	18.2	25.7	27.1
		author . .	—1.8	—5.4	— 9.1	— 9.6	—8.1	—6.5	—2.3	3.8	9.9	17.3	23.1	23.6
		NA	—0.7	—5.1	— 9.5	—10.2	—7.9	—5.1	—1.0	4.1	10.1	17.5	22.4	23.7

1.3, 1.1 and 1.1 m/s. The conclusion is that in this case no relationship between the standard deviation of the geostrophic wind on the one hand and R and the geostrophic wind speed on the other could be found. For an area containing about ten grid points selected from the region between Newfoundland and the British Isles and at the lower part of the grid the standard deviation is slightly larger than 1 m/s, originating from differences between the opinion of 26 analysts about the most probable solution of the analysis problem. Notice, however, that the analysis was rather simple in this case.

Final discussion

Part of the discussion in this paper has been devoted to the question of the variability of the height of the 500 mb surface as originating from the personal factor in upper air analysis. Furthermore, the influence of this variability on certain verification parameters has been investigated. It must not be forgotten, however, that the height values as given by the radiosoundings have errors too. Take e.g. a grid point near a weather ship. This point may have a standard deviation (σ_h) of 6 gdm. The standard deviation for the radiosounding itself is about 20 gdm. We notice that the

Table VI. Geostrophic wind speed for certain selected grid points (for coordinates see fig. 1). Values of maximum, minimum, mean and range are given for 26 analyses. Values of σ_g, R, σ_h and range of the height of the 500 mb surface as calculated from the 26 analyses are also given. Wind speed in m/s

Coord	x =	13	12	13	17	10	7	9	2	7	2	11	13	10	4
	y =	5	10	13	13	19	21	22	5	3	13	13	20	25	27
max.		16.8	24.3	39.0	24.0	23.8	26.7	14.8	29.3	33.1	9.0	33.7	48.8	28.7	14.4
min.		10.8	13.5	29.1	16.3	17.6	22.8	9.0	13.1	21.9	0.8	26.8	33.3	19.5	4.4
mean		14.0	17.3	32.7	19.7	20.0	24.8	12.1	20.7	28.5	3.7	30.3	41.5	23.6	8.3
range		6.0	10.8	9.9	7.7	6.2	3.9	5.8	16.2	11.2	8.2	6.9	15.5	9.2	10.0
routine		10.3	18.9	35.3	16.9	16.4	23.6	10.2	14.0	22.0	3.3	30.5	37.7	24.7	2.3
author		15.6	18.9	28.7	22.6	20.3	24.5	10.7	21.6	26.2	2.0	29.5	38.1	27.8	5.9
NA		15.1	17.5	31.2	19.5	19.2	24.7	14.8	19.2	26.3	4.0	28.6	35.9	24.0	8.3
σ_g		1.8	2.8	3.0	2.0	1.5	1.2	1.5	4.0	2.8	2.0	2.2	3.2	2.5	2.3
R		5	12	5	5	4	4	2	4	32	46	25	20	30	5
σ_h		2.7	6.8	6.2	5.8	3.4	4.1	4.7	43.8	36.4	43.1	25.9	24.6	28.2	32.0
$(\text{range})_h$		20	35	35	30	10	15	15	135	125	185	100	70	115	100

variability due to the personal effect in the analysis gives a standard deviation that is not negligible compared to the standard deviation of the individual sounding. If there is any doubt as to the reliability of the height value given by a sounding the standard deviation (σ_h) may be as large as 45 gdm. These additional errors in the analysis must at times have a serious repercussion on a numerical forecast based on that analysis. This effect has been discussed by NEWTON (1954) who gives a drastic example of what may happen if a reanalysis is made. BEST (1956) has discussed for three cases the differences between numerical forecasts made from different analyses. It would have been very interesting to make 29 different forecasts based on the 29 analyses discussed above. Unfortunately, this is not possible because of the great cost involved. The author hopes, however, to be able to make a numerical forecast from a small selection of the analyses and to report the result in a later paper.

The error in the determination of the wind is rather small. A reasonable value for radar wind and radio wind equipments seems to be of the order 2 m/s at 500 mb. It is not likely that this error has any noticeable influence on the analysis of the gradients, and the differences in the 26 analyses discussed in this paper may safely be said to be rather independent of the error in the wind observations. The same goes for the variability of the wind due to large eddies. Furthermore, the winds are plotted rounded off to the nearest five knots.

When using the isohypse field to estimate the geostrophic wind as is done in the forecast technique used by the group at the University at Stockholm, some errors creep into the picture. Firstly, the standard deviation of the radiosounding introduces an error in the geostrophic wind field. This effect has been discussed among others by HOVMÖLLER (1952). The standard deviation arising from the personal factor in chart analysis has a similar effect. Let us assume that we want to estimate the geostrophic wind with the aid of four grid points, e.g. over a distance of about 600 km. When we want to estimate the geostrophic wind at a point A (see fig. 1) the grid points B–D and C–E, respectively, are used simultaneously. The instrument standard deviation (σ_i) is supposed to be 20 gdm. The standard deviation of the geostrophic wind (σ_g) is

$$\sigma_g = \frac{g\sigma}{f \cdot \Delta n}$$

where g is the acceleration of gravity, f the coriolis parameter $\simeq 10^{-4}$ sec^{-1} and Δn the distance between grid points $\simeq$ 300 km. The value of σ is determined by the formula $\sigma^2 = \sigma_i^2 + \sigma_d^2$ provided that the instrument error (σ_i) and the standard deviation (σ_d) of the difference between the height values at the grid points B—D and C—E, respectively, are uncorrelated. This is generally not the case. If a radiosounding shows a value that differs markedly from its neighbouring stations, then different analysts have to make an individual interpretation of the height value, thus creating an increased value of σ_h. The effect ought to be more marked when the error is large. The value used here (20 gdm) is a moderate one and for that reason the simple equation may be used in order to make it possible to arrive at a reasonable estimate of σ_g.

The value of σ_d is a function of σ_h at the grid points and of the correlation coefficient between the height values. Giving σ_d the values 0, 15, and 30 gdm, respectively, we obtain the σ_g-values 6.5, 8, and 12 m/s, respectively. A large value of σ_d has the same effect as the instrument error, thus increasing the uncertainty of the geostrophic wind considerably.

The value of σ_d is as already mentioned a function of the standard deviation of the height values at the grid points (σ_h). On the other hand a definite relationship exists between R and σ_h. With a large material at hand it would be possible to arrive at representative values of R, σ_h and the correlation coefficient between the height values at the grid points used for the calculation of the geostrophic wind. Then, the value of σ_g may be determined as a function of R. This is interesting to note because it gives us a possibility to determine that value of R that corresponds to a predetermined value of σ_g. In other words, by making a series of investigations along the lines indicated here one might be able to determine the radiosonde station network necessary to make a certain predetermined accuracy of the analysis possible.

The orientation of the grid may have an

effect of the calculation, but nothing is known by the author about the magnitude of the variation thus introduced.

Finally the difference method introduces an error. Instead of calculating derivatives, we are calculating ratios of finite differences. The result is a smoothing of the field. In the numerical forecast techniques as used for the time being this effect is mostly a wanted one. As a matter of fact the calculations quite often start with a smoothing of the initial field. It must be emphasized, however, that not enough is known about the error introduced by using finite differences and about the exact repercussion of it on the numerical calculations.

Taking all the above effects into account it is obvious that the geostrophic approximation used for forecasting purposes is affected by a rather large total standard deviation.

Not much is known of the overall effect of this error. One way to solve this question is to investigate a number of well-defined streamline patterns and using both a large number of analysts to get conventional analysis and a high-speed computer to calculate a variety of parameters. In this paper, however, only the error of the estimation of the geostrophic wind field has been discussed. In a later paper the variability of the vorticity and the advection of vorticity will be discussed to some extent. The much more difficult question of the error in the geostrophic approximation itself has still to be solved.

Acknowledgement

Thanks are due to Dr A. Nyberg, Director of the Swedish Met. and Hydr. Institute, Stockholm, and to Dr B. Bolin, University of Stockholm, for the great interest they have shown in this work. The author is also indebted to Mr B. Döös, University of Stockholm, and to Mr T. Duvedal, SMHI, Stockholm, for their very great help with the calculations made on the high-speed computer (BESK).

REFERENCES

Berggren, R., 1952: The distribution of temperature and wind connected with active tropical air in the higher troposphere and some remarks concerning clear air turbulence of high altitude. *Tellus*, **4**, No. 1, pp. 43—53.

Bergthorson et al., 1955: Routine forecasting with the barotropic model. *Tellus*, **7**; No. 2, pp. 272—274.

Bergthorson, P., and Döös, B., 1955: Numerical weather map analysis. *Tellus*, **7** No. 3, pp. 329—340.

Best, W. H., 1956: Differences in numerical prognoses resulting from differences in analyses. *Tellus*, **8**, No. 3, pp. 351—356.

Bolin, B., 1955: Numerical forecasting with the barotropic model. *Tellus*, **7**, No. 1, pp. 27—49.

Charney, J., 1949: On the physical basis for numerical prediction of large-scale motions in the atmosphere. *J. of Met.*, **6**, pp. 371—385.

Charney, J., and Phillips, N., 1953: Numerical integration of the quasi-geostrophic equations of motion for barotropic and simple baroclinic flows. *J. of Met.*, **10**, pp. 71—99.

Döös, B. R., 1956: Automation of 500 mb forecast through successive numerical map analyses. *Tellus*, **8**, No 1, pp. 76—81.

Döös, B. R., and Eaton, Mc., 1957: Upper-air analysis over ocean areas. *Tellus*, **9**, No. 2, pp. 184—194.

Gates, W. L., Pocinski, L. S., and Jenkins, C. F., 1956: Results of numerical forecasting with the barotropic and thermotropic atmospheric models. *Geophys. Res.* Pap. No. 46.

Hovmöller, E., 1952: A comparison between observed and computed winds with respect to their applicability for vorticity computations. *Tellus*, **4**, No. 2, pp. 126—134.

Newton, C. W., 1954: Analysis and data problems in relation to numerical prediction. *Bull. Am. Met. Soc.*, **35**, No. 7, pp. 287—294.

Nyberg, A., 1952: On the comparison of radiosonde data in Payerne, May 1950. *Sveriges meteorologiska och hydrologiska institut*, Medd. serie B, Nr 9, 22 pp.

Sawyer, J. S., 1950: Equivalent headwinds, application of upper-wind statistics to air-route planning. *Met. Reports*, No. 6, Met. Office, London.

Sawyer, J. S., and Bushby, F. H., 1953: A baroclinic model atmosphere suitable for numerical integration. *Journ. of Met.*, **10**, No. 1, pp. 54—59.

Staff Members, Inst. of Met., Univ. of Stockholm, 1954: Results of forecasting with the barotropic model on an electronic computer (BESK). *Tellus*, **6**, No. 2, pp. 139—149.

Noctilucent Clouds

By F. H. LUDLAM, Imperial College, London

(Manuscript received December 12, 1956, revised March 11, 1957)

Abstract

Knowledge of noctilucent clouds is reviewed, and fresh observations made in Sweden are described. The conditions under which the clouds may be seen are calculated and compared with observations; it appears that the clouds are formed only in the summer in high latitudes. The sizes and concentrations of the cloud particles are inferred, and the probsem of their composition and origin is discussed. The clouds are not likely to be formed by condensation of water vapour, but are more probably dust clouds derived from great eruptions, very large meteorites, and from interplanetary space. Evidence of dust in the stratosphere is assembled, and the processes by which it appears in visible clouds are discussed.

1. Introduction; the properties of noctilucent clouds

In high latitudes during summer months tenuous high clouds occasionally become visible about an hour after sunset. Only clouds high in the stratosphere can remain in sunshine so long after the sun has set on the ground.

Such clouds, called 'luminous night-clouds' or *noctilucent clouds*, were first studied systematically by Jesse, in Germany, and by Tseraskii, in Russia, in 1885 and later years. It is probable that similar clouds had been observed previously, but the published accounts are not sufficiently definite for this to be certain (see ARAGO, 1854, and SCULTETUS, 1949, who discusses an observation by Lavoisier). There are numerous reports of bright noctilucent clouds in 1885, 1886 and 1887, but in succeeding years there were fewer observations and the displays became weaker, and since 1894 the clouds have been observed only intermittently. A valuable review of published reports was made by VESTINE (1934), and more recently STØRMER (1933, 1935) and PATON (1954) have on several occasions observed and measured the height of the clouds.

The problem of the nature of the clouds is still unsolved. In this paper some new observations are described, previous reports are summarized, and theories of the composition and formation of the clouds are discussed, with the aim of stimulating more widespread and thorough observations during the approaching International Geophysical Year.

a) *The form and colour of noctilucent clouds*

In general appearance noctilucent clouds resemble a thin, decayed cirrostratus, without any of the fallstreak forms characteristic of many ice clouds (Figs. 1—3). They tend to occur in extensive sheets, although the part of the sky in which they can be seen is often restricted to a narrow segment above the horizon. The most careful classification of the forms of noctilucent clouds is that of GRISHIN (1955); he distinguishes the following:

1. The *nebula*, or flare. Between the more prominent details of other forms there is frequently a nebulous, very tenuous and structureless veil, shining faintly with a delicately white or bluish light. According to Grishin the flare often can be detected by an experienced observer as a slight luminescence of the twilight sky, and may appear a half to one hour before clouds with a definite structure.

2. *Bands*. Diffuse-edged, long streaks, known as bands, are more or less straight or gently

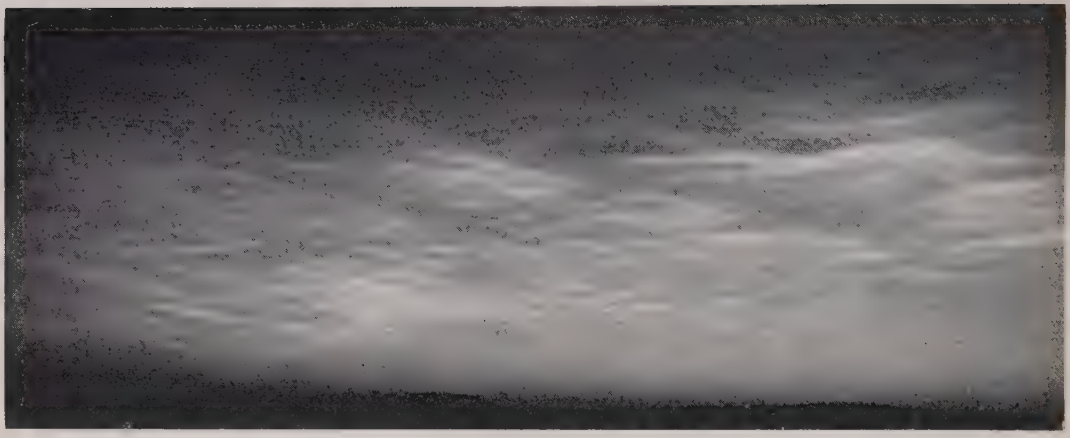


Fig. 1. Noctilucent clouds in the northern sky, Torsta, 13 August 1955, 0025 hours. A dark "shadow" can be seen above a bright detail in the upper right of the picture.

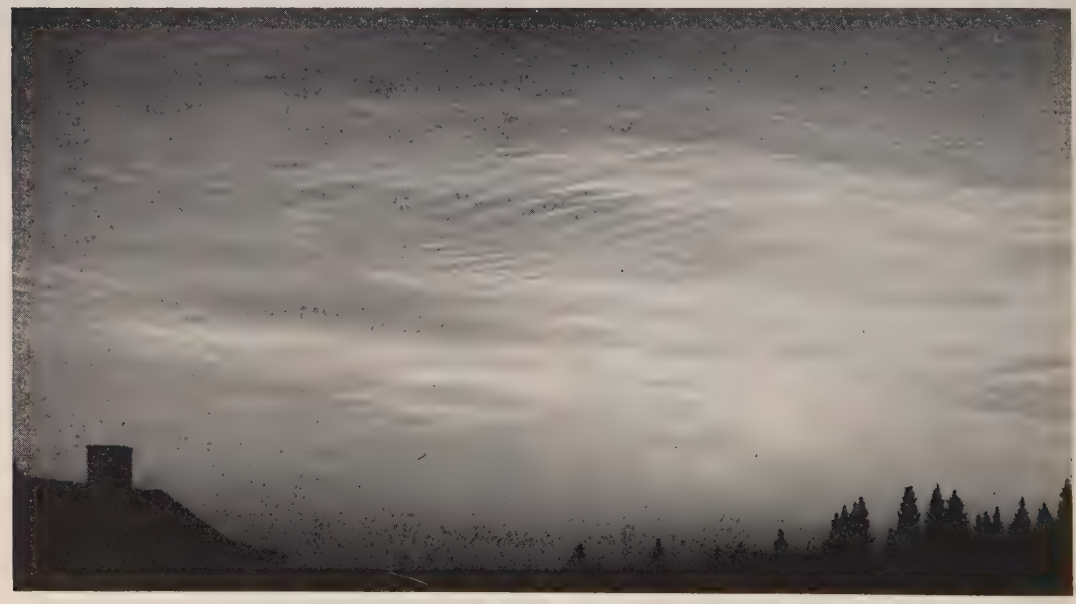


Fig. 2. Noctilucent cloud billows, Torsta, 13 August 1955, 0155 hours (photograph by F. Singleton).

curved, and sometimes hundreds of km long. They occur in extensive groups, the individuals of which are roughly parallel. They are persistent, with inappreciable change of detail. Smaller *streaks*, with twists or bends, may lie across the bands or branch out from them.

3. *Billows.* When the noctilucent clouds extend above low elevations, groups of rather sharply-defined billows may be seen. The distance separating pairs of billows is about 10 km. The billows tend to lie across the direction of the bands, but their alignment may differ noticeably in neighbouring parts of the sky. When the billows lie across a band part of the material of the band becomes gathered into the billow crests; at other times billows form within the flare. In contrast to the bands, the small streaks and billows may change their form and arrangement, or appear and disappear, within several minutes.

The colour of the clouds is usually white; at high elevations they are often silvery, having a slight bluish tinge, while near the horizon they become yellowish, or even orange. Sometimes more closely-spaced *serrations* are seen, which to the unaided eye present an almost continuous cloud. Grishin also distinguishes individual wave-like curves or distortions in parts of noctilucent clouds, and

4. *Eddies*, usually of slight curvature, which may be seen in bands, serrations, and sometimes in flares. The eddies extend over very variable arcs, even to complete rings or vortices with dark centres.

Grishin's classification was developed during observations made in the years 1948 to 1953; in the last year a time-lapse film of noctilucent clouds was made, which clearly showed the wave-like form and evolution of many of the structural details. No satisfactory impression of the nature of the internal motion and transformations shown by these clouds can be gained except by the help of such films.

According to Helmholtz, Størmer, and Astapovich (see KHVOSTIKOV, 1952), their spectrum is similar to that of daylight; the Fraunhofer lines are clearly present, and there are no emission lines. The light from the clouds is evidently simply scattered sunlight, but with some excess of blue compared with red. PATON (1954) once observed a slow change of colour in portions of a noctilucent cloud from a vivid blue to white.

Occasionally the clouds are strikingly bright. In 1885 and 1886 they were bright enough at times to cast shadows and to allow the reading of ordinary print outdoors. A report in 1908 (WHIPPLE, 1934) speaks of 'an extraordinarily strong light, magical and very imposing in the fair summer night'. PATON (1951) mentions a display during which notes could be written without artificial light. Nevertheless the clouds are usually very tenuous, causing hardly any dimming of stars seen through them. During moderate displays in Sweden the clouds were photographed on fine-grain panchromatic film (speed 28° Sch.) with exposures varying from about 1 sec, with a solar depression of 6°, to as much as 100—500 sec, with a solar depression of 11—12.5°, at an aperture of f 5.6. The clouds are seen clearly through blue filters, but fade in red filters, and the latter, so useful in the photography of cirrus, should not be used when photographs are taken of noctilucent clouds.

The light of the noctilucent clouds is quite strongly polarized, in the same sense as the light of the sky. This property is sometimes useful in distinguishing faint noctilucent clouds from cirrus feebly illuminated by scattering from the twilight sky.

b) *The occurrence of noctilucent clouds*

According to previous accounts noctilucent clouds are observed only at places lying between latitudes of about 45° and 62° N (see, e.g. PATON, 1954). However, they were observed on ten nights during the months July and August, 1954 and 1955, at the field-station of the International Institute of Meteorology (Stockholm), which is at Torsta, near Östersund in central Sweden, and in latitude 63°15′ N. On some occasions the clouds were seen in the northern sky down to elevations of 1 to 2° above the horizon, corresponding to distances of 800 to 900 km. These clouds therefore extended northwards at least into the vicinity of North Cape, in latitude 71° N (only in the clean air so often enjoyed in this part of the world could it be expected that clouds would be visible at such a distance). The previous lack of reports from north of 62° N is therefore probably due to the scarcity there of keen observers.

It may equally be wondered if the complete lack of reports during some years reliably indicates the absence of the clouds. The decrease in the number of accounts after 1887 may have been due partly to waning interest. PATON (1954) remarks that the records from Ben Nevis Observatory, where hourly observations were made from 1883 to 1904, show that noctilucent clouds were frequently seen and noted as 'pearly cirrus', 'shining with a silverly blue light'. Paton himself, observing during the period 1939 to 1955, saw noctilucent clouds at least once each year, except in 1942 and 1944 (private communication). If these observations are added to those of Størmer, Malzew (quoted by STØRMER, 1935), and SPANGENBERG (1949), we find that only in these two years are reports missing in the years since 1932. It seems likely, therefore, that the clouds occur at least somewhere in high latitudes every summer, that often they are not seen for want of suitable weather conditions or interested observers, and

that occasionally they are abnormally widespread and intense, compelling attention. The exceptional periods seem to be 1885—1892, 1908—1911, 1932—1935, and, possibly, 1953—1955.

The clouds have been seen in the northern hemisphere from late May to the middle of August, and, according to Vestine (who does not, however, quote references) in the southern hemisphere in the period of the year about six months later.

Fig. 4, from Vestine's review, shows that the clouds have been observed more frequently after than before the summer solstice.

During strong displays the clouds have appeared as early as 15 minutes after sunset over a large part of the sky, extending to near the SE horizon. As the sun sinks lower the illuminated area of the clouds recedes to a zone above the twilight arch, on either side of the vertical through the sun. This zone moves with the sun's vertical into the northern sky, where at midnight the upper edge of the illuminated area has reached a minimum elevation, and towards sunrise into the northeast before extending again across the zenith and fading away about half an hour before sunrise. In less striking displays the clouds appear in the northwest about an hour after sunset, and fade in the northeast about the same time before sunrise. Only rarely have the clouds been seen at elevations higher than 10°.

Except in very intense displays, the clouds have been seen more frequently after than before midnight, even though the number of observers is likely to have been less at the later hours. In the period 1889—1894 the clouds were observed on 6 occasions before, and on 33 occasions after midnight.

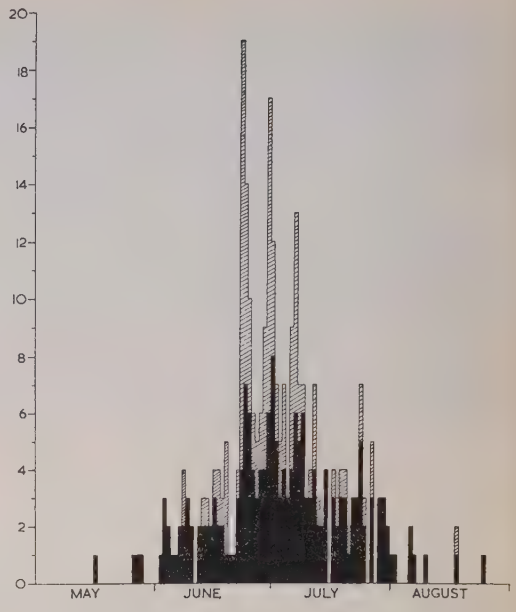


Fig. 4. Total number of times noctilucent clouds have occurred on a given date, 1885—1933 (solid histogram); total number of reports of noctilucent clouds on a given date (hatched histogram). (After Vestine, 1934.)

c) *The height of noctilucent clouds*

Tseraskii in 1895, and Pokrovskii in 1897 (see KHVOSTIKOV, 1952), JESSE (1896), STØRMER (1935) and PATON (1949) have measured the height of noctilucent clouds by photography from the ends of long base-lines. Tseraskii obtained an approximate height of 79 km, and Pokrovskii a value of 82 km. Jesse used a base-line 35 km long; in the summer of 1889 he determined the height of 108 cloud details identified on simultaneously exposed pairs of plates, and obtained an average of 82.8 km. When these were combined with another 179 measurements made in the following summer, the mean became 82.1 km; the individual values varied from 79 to 90 km.

Størmer used his network of auroral cameras, which provides base-lines of 47, 65 and 105 km. On two occasions in 1932 he measured mean heights of 81.8 km (18 measurements) and 81.1 km (19 measurements). On another occasion in 1934 he obtained a mean height of 82.2 km from 41 measurements which varied from 78 to 85 km.

Paton also used auroral cameras on an occasion in 1949, with a base-line 27.6 km long, which, however, was unfavourably orientated with respect to the clouds, lying almost N—S; he measured a height of about 89 km.

The measurements are thus very consistent. Estimates based on the angular elevation of the nearest discernible details, and the assumption that they are lit by rays which graze the earth, give smaller height varying between about 30 and 70 km.

d) *The velocity of noctilucent clouds*

Jesse made a number of measurements of the movement of noctilucent clouds and found it

a.

b.

Fig. 3. Noctilucent clouds in NNE, 9 August 1955, 2340 hours (a), and 10 August 1955, 0010 hours (b).

to be almost always from some point between N and E, and to be rapid, with speeds of up to 100—200 m/sec (ARCHENHOLD, 1928). In 1932 Størmer observed the clouds to be moving from NNE at 44—55 m/sec, and in 1935 estimated a movement from E at about 80 m/sec. STØRMER (1935) quotes a measurement made in Russia in 1925 of a motion from 013° at 229 m/sec. KHVOSTIKOV (1952) mentions measurements in Russia of speeds varying from 50 to 135 m/sec.

According to Jesse's data the motion before midnight is usually from NE, while after midnight it tends to be from ENE. STØRMER (1933) suggests that the changes in velocity may occur near the earth's shadow.

2. Summary of observations at Torsta, 1954 and 1955

In the following summary of observations made at Torsta, near Östersund, a note is included of occasions when the clouds were not seen although sought under apparently favourable viewing conditions. The times quoted are in Swedish clock time (G.M.T. plus 1 hour), which within a few minutes is also the local time. The sun's declination is given in brackets following each date of observation. Cloud velocities are derived from theodolite observations, on the assumption that the cloud height is 80 km. Figures following bearings refer to elevations above the horizon.

18—19 July 1954 (21.0°)

The clouds were first noticed in the NE, 20° at 2322, and later appeared in various places between NNW and E, up to 30°. Billows were aligned E—W. About 0040 four successive measurements gave the following velocities: 003° 55 m/sec, 021° 70 m/sec, 009° 66 m/sec, 011° 65 m/sec.

The clouds were also observed in Stockholm from 2400 to 0100, between N and NNE (fig. 5).

20—21 July 1954 (20.8°)

A good display occurred after 2300, covering practically the whole sky north of the zenith, mainly as very long irregular streaks aligned N—S. Some billows in the east had the same alignment. At 0012 the clouds were visible in the NNW down to 6°, where they gleamed

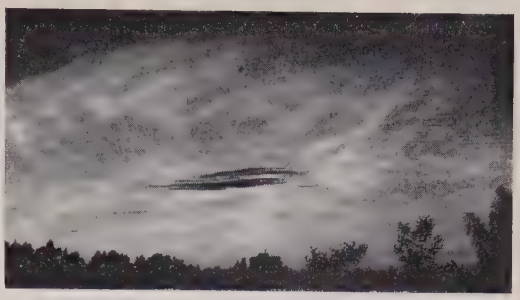


Fig. 5. Noctilucent clouds in NNE, from Lidingö, Stockholm, 19 July 1954, about 0030 hours (photograph by Hans Höfer).

silver behind a rayed purple light extending up to about 10°; one measurement gave a velocity of 345° 86 m/sec. At 0035 the clouds extended to 18° above the SSW horizon; the display faded at 0200.

24 July 1954 (20.0°)

The clouds did not occur before 2330.

29—30 July 1954 (18.9°)

Some clouds were barely visible beyond stratocumulus, just E of N, 5°, at 2345. Later long streaks and billows appeared in the NE, about 10°.

16—17 August 1954 (13.9°)

Noctilucent clouds appeared at 2300 in an arch extending from WNW to NNE, with minimum and maximum elevations of 2.2 and 7° (where they faded sharply into a deep blue-black sky) in the NNW at 2335, and were still visible at 0010. One measurement at 2342 gave a velocity of 053° 55 m/sec.

17, 19 and 22 August 1954 (13.6, 13.0, 12.0°)

The clouds were not present before 2400.

18 August, 1954 (13.3°)

The clouds were not present before 2300.

29 August 1954 (9.6°)

The clouds were not present before 2200.

18—19 July 1955 (21.0°)

Between 2345 and 0035 faint streaks were visible near the zenith, down to 40° in the S and down to 10° between W and NNE. No appreciable movement of the streaks could be

seen over the open sights of the theodolite during a ten-minute period.

23—24 July 1955 (20.0°)

At 2340 faint streaks lying N—S appeared in NE and at about 30° in the SE a faint streak was aligned E—W. At 2400 faint patches with billows appeared in SE, 60°; motion 355° 70 m/sec. At 0015 further streaks appeared in NW, 25° and ENE, 15°. Another measurement of the motion, more reliable than the first, gave 022° 80 m/sec. At 0110 the display had faded, but two long faint streaks were still visible in the E, 25°.

25—26 and 26—27 July 1955 (19.7 and 19.5°)

No noctilucent clouds before 0030.

27—28 July 1955 (19.2°)

At 2250 a thin streak appeared in N-NNE, 3°; soon another appeared at 6°, and a patch of billows formed just below 3°. At 2325 the clouds had extended up to 7.5°; they gave a strong impression of forming and dissolving with a life of about 20 min. From 2345 to 0040 there were only traces, from NNW to NNE, 3.5 to 4.6°.

29—30 July 1955 (18.8°)

No clouds present before 0030.

31 July—1 August 1955 (18.3°)

At 2245 pale streaks became visible in NNE, 7°; gradually they became brighter and more extensive; the longest streaks were aligned towards ENE, gradually bending towards NNW at low elevations, with many shorter streaks and some billows lying across their length. By 2330 the clouds extended from NNE to ENE, 1.7 to 10°. Immediately above bright streaks the sky was darker, as though they cast a shadow. At 2350 the motion was measured as 344° 36 m/sec. At 0019 and 0023 two further measurements in the same part of the sky, apparently no less reliable, gave 238° 45 m/sec and 243° 38 m/sec. The alignment of a billow was measured to be along 293°. After midnight the display became weaker, and only traces remained at 0040.

2 August 1955 (17.8°)

Clouds not present before 2315.

9—10 August 1955 (15.8°)

Noctilucent clouds appeared in the N at 2330, and by 2350 extended from NNE—ENE, 2.0 to 5.0°. Two measurements of the motion gave 325° 13 m/sec (at 2400) and 341° 16 m/sec (at 0025). By 0040 the clouds extended from 3.2 to 13.4°, but were fainter; throughout this period they were very sharply limited in the NNE. At 0145 traces remained in ENE, 11 and 6°. The clouds are illustrated in Fig. 3.

12—13 August 1955 (15.0°)

Clouds first observed at 2330; at midnight they extended from NNE to ENE, up to 4.3°; half an hour later they had spread into the N, with traces also in the NW, and the upper margin had increased to 6.2°. The area of clouds continued to shift westwards and the upper margin to rise; at 0135 the clouds extended between NNW and NNE and up to 12.8°. There were many groups of beautiful billows (Figs. 1, 2), and again marked 'shadows' occurred immediately above some of the brighter streaks (Fig. 1). After 0200 the clouds faded and shrank in extent, only traces in the NW remaining at 0225. Measurements of the motion gave the following results:
0100: 031° 36 m/sec; 0130: 052° 26 m/sec and 049° 60 m/sec; 0150: 054° 90 m/sec.

The acceleration seemed real, for in the later stages the motion in the field of the theodolite was obviously more rapid than previously.

21—22 August 1955 (12.1°)

Noctilucent clouds absent.

23, 24, 25, 26 August 1955 (11.4 to 10.4°)

Noctilucent clouds absent before 2330.

General remarks

Altogether during these two summers there were observed 2 faint, 3 moderate and 5 bright displays. It is remarkable that between successive displays the clouds are sometimes not to be seen under apparently perfect viewing conditions. Possibly on some of these occasions a watch, although maintained until about midnight, was not kept long enough, for in accordance with previous experience there is a clear tendency for the displays to be more marked after than before midnight. The clouds also are usually visible for a longer period after than before midnight.

It is noticeable that the clouds occur predominantly in the part of the sky which is east of N, and even for them to first appear there before midnight, when of course the sun lies west of N. On some occasions the clouds are clearly limited in extent, their boundaries not being determined simply by the border of the earth's shadow.

It is not easy to make a reliable measurement of the cloud motion by theodolite, because of the difficulty of finding a suitable detail which retains its identity and sharpness for more than a minute or two. Some of the prominent details, for example the bright places where two streaks cross or fuse, may not move with the wind velocity.

3. Conditions under which noctilucent clouds may be observed

It has been supposed that observations of the clouds are restricted to the summer months in middle and high latitudes because only then do the geometrical conditions allow the clouds to remain in sunshine for substantial periods after sunset on the ground. In low latitudes throughout the year the sun sinks rapidly below the horizon, greatly restricting these periods in which the clouds are suitably illuminated. The variation of viewing conditions throughout the seasons and over the earth can be calculated under some assumptions, as described in the following paragraphs.

An observer at O (Fig. 6) is considered able to see any clouds lying at the 80 km level over an arc extending in the vertical through the sun between angular elevations a, b, when the sun's apparent depression below the horizon is α. The sunlight grazes the earth at the point G,

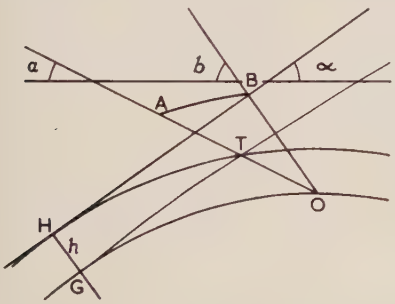


Fig. 6. Geometrical conditions under which noctilucent clouds may be seen.

but it is supposed that the light is much attenuated in a layer extending above the ground to a height h, so that only rays above that which grazes this layer at H illuminate the clouds sufficiently for them to become visible. On the other hand, it is assumed that the clouds cannot be seen beyond the point A because in the line of sight a part of the main scattering layer is still in direct sunshine, and the cloud details are swamped in the flood of scattered light which produces the twilight arch.

The angles a and b are functions of the sun's depression α, and of the heights assumed for the clouds and the top of the main scattering layer. They may be determined by calculating the paths through the atmosphere of the rays along OA and OB, and of the rays GT and HB which graze the earth's surface and the main scattering layer. Because these rays have long paths through the atmosphere it is necessary to consider the earth's curvature, and at the same time normal atmospheric refraction can be taken into account (although it is relatively unimportant).

It is desirable to determine a value of h from observations of the upper and lower boundaries of noctilucent clouds. Størmer (1933, 1935) has determined from his observations the height of the earth's shadow at the position B of the upper edge of the clouds. Although he neglected refraction he in this way obtains a close approximation to h, with results varying from 29 to about 50 km. Only the minimum values are likely to be significant, because on some occasions there may be no clouds extending into the position B; nevertheless in six measurements made on each of two occasions Størmer obtains values which rather consistently lie between 30 and 45 km. This astonishing result implies either that there is very substantial attenuation of sunlight even well in the stratosphere, or that the clouds do not quite extend to the position of the earth's shadow.

I have calculated values of h for the rays which illuminated the nearest and furthermost details of the clouds observed at Torsta, assuming the cloud height to be 80 km and taking into account refraction, with the results shown in Table 1. These confirm Størmer's values, and suggest that only rarely are the nearest details in a position corresponding to a value of h smaller than 30 km. The values of h

Table 1. Least height h above the ground of rays illuminating nearest and farthest details of noctilucent clouds

Date	Time	Sun's depression, deg.	Azimuth and Elevation of clouds, deg.	h, km
Nearest details:				
18.7.54	2400	5.7	360,30	54
21.7.54	0035	6.0	90	45
16.8.54	2335	12.8	330,07	12
19.7.55	0005	5.7	180,40	43
23.7.55	2400	6.8	90	34
27.7.55	2325	7.5	360,07.5	69
31.7.55	2330	8.5	360,10	55
9.8.55	2350	10.8	040,05	37
10.8.55	0025	11.0	031,06.3	34
	0040	10.7[1]	035,13.4	10
13.8.55	0040	11.4	033,06.3	32
	0115	10.6	015,09.9	30
	0140	9.7	007,12.8	29
	0215	8.0	353,10.0	45
Farthest details:				
19.7.54	0025	5.7	360,04	12
21.7.54	0015	6.0	340,06	16
16.8.54	2335	12.8	320,02.2	55
19.7.55	0005	5.7	360,10	16
27.7.55	2330	7.5	360,03.0	18
31.7.55	2330	8.5	020,01.7	23
9.8.55	2350	10.8	023,03.3	38
10.8.55	0025	11.0	023,02.9	37
	0040	10.7	024,03.2	36
13.8.55	0040	11.4	006,01.2	32
	0115	10.6	360,01.7	31
	0140	9.7	007,01.8	26
	0215	8.0	344,01.6	25
			008,04.7	25

[1] (very faint)

corresponding to the farthermost visible details seem significantly less, and bearing in mind that the minimum values are likely to be more significant, I have used a value of 16 km in the following calculations. It must be supposed, however, that the appropriate value may often be twice as much, with the effect of restricting the visible area of the clouds. It should be noted that the scattering of indirect light which limits the visibility of the clouds is that which occurs in the lower stratosphere, and that in the troposphere only attenuation by concentrations of haze or obscuration by clouds are likely to interfere with observations.

Fig. 7 shows the relations between the sun's apparent depression α and the angles a (for a value of $h=16$ km) and b ($h=30$ km), and Fig. 8 shows the relation between the distance of noctilucent clouds at 80 km and their angular elevation, used in the derivation of Fig. 7. This diagram shows that the clouds are likely to be visible only when the sun lies between about 5° and 16° below the horizon.

4. Variation of conditions of viewing according to latitude, season, and time of day

The true depression $-\gamma$ of the sun is a function of latitude, ϕ, sun's declination δ, and apparent solar time, τ, according to the equation:

$$\sin \gamma = \sin \phi \sin \delta + \cos \phi \cos \delta \cos \tau \qquad (1)$$

For a given latitude a diagram can be constructed with apparent solar time as ordinate

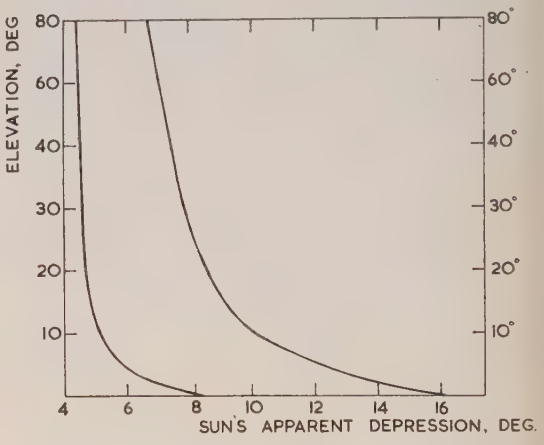


Fig. 7. Angular elevation of upper and lower borders of region, above sun's azimuth, in which noctilucent clouds may be seen, as a function of the apparent depression of the sun.

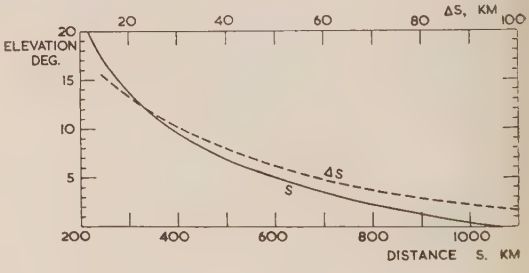


Fig. 8. Distance S of noctilucent clouds 80 km above the level of the observer, as a function of elevation above the horizon, and the change ΔS in S due to a change of 1° in the elevation.

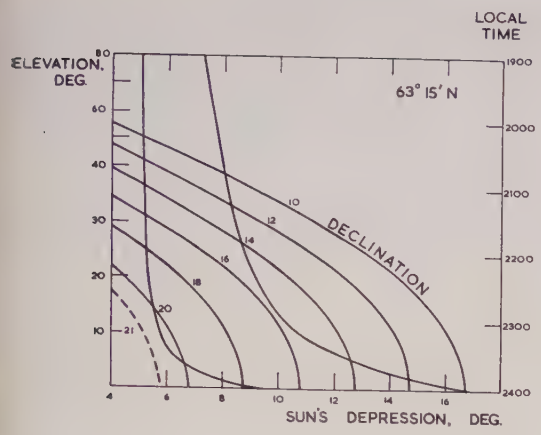


Fig. 9. Diagram for finding upper and lower border of noctilucent clouds on sun's azimuth for an observer in latitude 63° 15′ N, as a function of the sun's true depression (assuming 0.6° of refraction at the horizon), or as a function of local time and sun's declination.

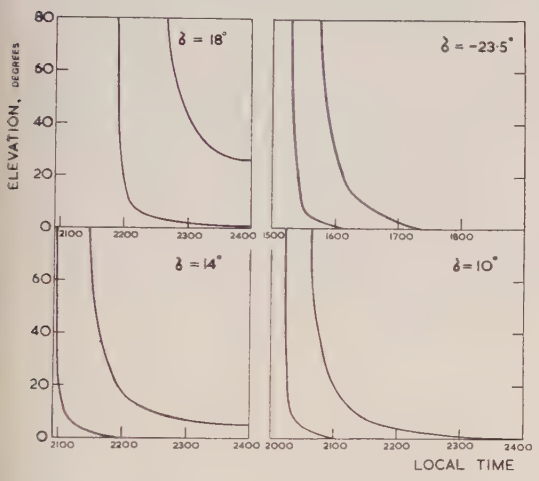


Fig. 10. Position of upper and lower borders of noctilucent clouds on the sun's azimuth for an observer in latitude 63° 15′ N, as a function of local time, for four times of the year.

and sun's true depression as abscissa, containing isopleths of declination. Such a diagram may have the curves of Fig. 7 superimposed, as for example in Fig. 9, since we can regard the true sun's depression to exceed the apparent by 0.6°, and it can then be used to find the elevations in the vertical through the sun where noctilucent clouds should be sought, as a function of season and apparent solar time. (The apparent solar time is found from the following relation: apparent solar time = = G. M. T. + (4 min. for each degree of longitude east of Greenwich)—(the equation of time). Values of δ and of the equation of time are tabulated for each day of the year in the Nautical Almanac.) The diagram is used by determining the point corresponding to a given declination and time; the vertical through this point intersects the *a*, *b*- curves in two points defining the lower and upper borders of the part of the sky above the sun in which noctilucent clouds may be seen.

The form of the *a*—*b* curves in Fig. 7 implies that in the vertical above the sun the upper and lower borders of the visible area of clouds pass rapidly through the zenith, but descend rather slowly as they approach the horizon. This is illustrated in Fig. 10, which shows the position of these borders with time of day for several seasons in latitude 63° 15′ N. These diagrams also show the most favourable season for viewing to be near mid-summer.

As a measure of an observer's opportunity of seeing noctilucent clouds it is reasonable to consider the total period during which a cloud at an elevation of 10° on the sun's azimuth is suitably illuminated. From Fig. 9 it is seen that during this period the true depression of the sun changes from 5.75 to 10.7°; the period can therefore be calculated from eq. (1) by inserting these values of $-\gamma$, for particular latitudes and seasons. In this way we derive Fig. 11, which shows the length of the viewing period (before midnight) throughout the year in latitude 63° 15′ N. Over most of the year the period is nearly an hour, rising sharply to a peak of nearly three hours at a declination of 15°, and abruptly falling to zero at declinations exceeding 21°, corresponding to one month before and after the summer solstice, when the sun even at midnight is less than 5.75° below the horizon. At rather lower latitudes mid-summer is the most favourable viewing season, as shown in Fig. 12. In this diagram the curves show the variation of the viewing period (before midnight) for two months after the summer solstice, in latitudes 63° 15′ N, 56° N, 50° N and 30° N. In the lower latitudes the period changes little throughout the year, and amounts to between a half and three quarters of an hour.

Such a period must be considered ample for an attentive observer to detect the clouds if they are present. Nevertheless according to

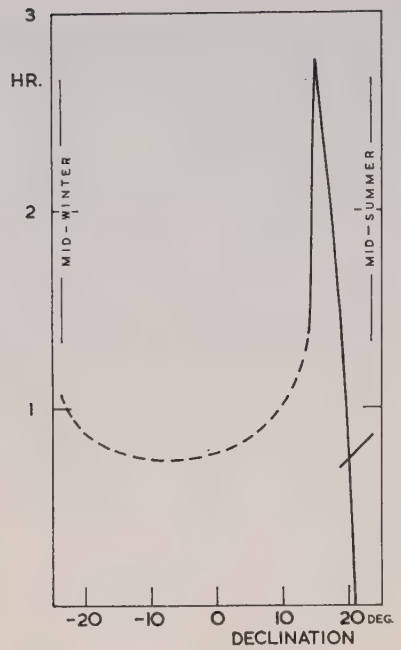


Fig. 11. Period before midnight during which noctilucent clouds on the sun's azimuth and at an elevation of 10° above the horizon are continuously visible, as a function of the sun's declination. The long curve is for an observer in latitude 63° 15′ N, and is drawn continuously over the season in which the clouds have been seen from Torsta. The short intersecting curve is the corresponding one for the latitude 50° N, and extends over the season in which the clouds have been reported from places between the latitudes 49° and 52° N.

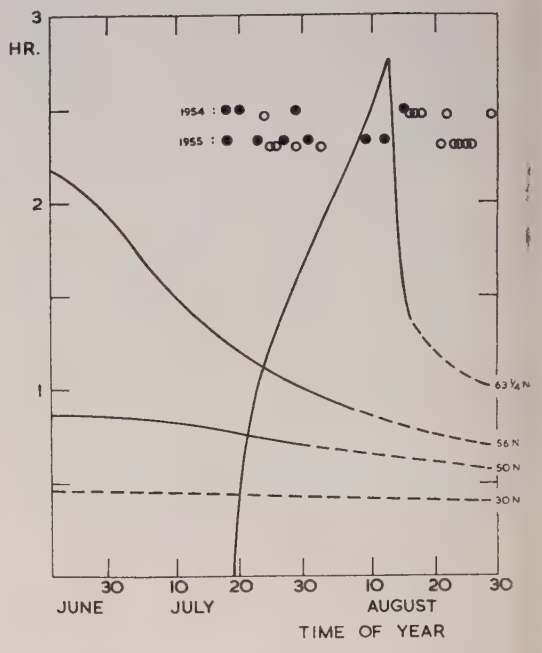


Fig. 12. Period before midnight during which noctilucent clouds on the sun's azimuth and at an elevation of 10° above the horizon are continuously visible, as a function of time of year. The continuous parts of the curves show the season during which the clouds have been observed at Torsta ($63^1/_4$° N), Edinburgh (56° N), and at places between 49 and 52° N (50° N). The circles in the upper part of the diagram show the occasions when the clouds have been seen at Torsta (blocked-in circles), or not seen under apparently favourable viewing conditions (open circles).

Vestine the clouds have never been reported from latitudes lower than 45°; apparently the reports from as far south as this were made in Russia on an exceptional occasion, that of the fall of the Great Siberian Meteorite, in 1908, which is further discussed below. If this occasion is excluded it seems that there are very few reports from south of 50° N, and if we consider also that the clouds seen in places near this latitude have rarely extended more than a few degrees above the northern horizon, then the presumption is strong that noctilucent clouds do not exist in latitudes lower than about 55°.

In Figs. 11 and 12 the curves are drawn as continuous lines over the season within which reports of the clouds have been made. The Torsta observations were used for the curve corresponding to latitude 63° 15′ N, Mr. Paton's (Edinburgh) observations for 56° N, and all those listed by Vestine from places between 49 and 52 1/2° N for the curve appropriate to 50° N. Evidently the observations of the clouds at Torsta and at Edinburgh cease when the viewing conditions, judged on this criterion, are still more favourable than ever occur in rather lower latitudes. At Torsta on clear days in mid-winter clouds 10° above the horizon would be visible for a whole hour in the morning and again in the afternoon, at times when they could hardly escape attention. It may also be significant that the clouds have been seen at Torsta rather later in the year than has ever been found at Edinburgh. On the basis of this evidence I conclude tentatively that the clouds, in addition to occurring mainly after rather than before mid-summer, are restricted in southern extent to about latitude 55° in July and to about 70° N in the latter part of August, and do not occur in these regions

in seasons other than the summer. These circumstances suggest some meteorological control of their appearance.

5. The size and concentration of the particles in noctilucent clouds

The noticeable degree of polarization in the light from noctilucent clouds indicates that

$$\pi r < 0.5\ \mu,$$

where r is the predominant radius of the particles in the clouds. On the other hand, the scattering of light by the particles cannot be that which occurs in the Rayleigh regime, for if so the clouds would appear blue. They are observed to be white, although sometimes tinged with blue. This shows that

$$\pi r > 0.025\ \mu,$$

and combining these indications we have

$$10^{-6} < r < 10^{-5}\ \text{cm}.$$

In Sweden it was found that when the sun was about 6° below the horizon, the exposure needed to photograph the clouds on fine-grain panchromatic film (speed 28° Sch.) was about 5 sec at f 5.6; if, however, the sun's depression was about 12°, the exposure needed was about 200 sec. It can be assumed, then, that an exposure of about 100 sec at this aperture is a measure of the brightness of typical clouds. This information allows the likely concentration of the cloud particles to be inferred, in the following manner.

It is assumed that the scattering particles are disposed in a layer of depth h, illuminated by the sun at an angle of incidence Θ. If a fraction $f(h)$, $\ll 1$, is scattered from a vertically incident solar beam, then a fraction $f(h)$ sec Θ is scattered from the beam incident at an angle Θ. Then the scattering fom a unit area of the cloud surface is independent of Θ and equal to $s \cdot f(h)$, where s is the flux/cm² in the solar beam. If the particles are small, half of the scattering is forward, and half backward.

If the cloud surface is assumed to scatter according to Lambert's law, its brightness B_c is independent of the direction of view and

$$B_c \sim s \cdot f(h)/2.$$

Over a surface of snow, or some other nearly perfect scatterer, which is illuminated vertically, the fraction of the incident beam scattered backwards is unity; thus if B_s is its brightness, $B_c \sim s$, and

$$B_c/B_s = f(h)/2.$$

We now suppose that the ratio of the brightness of these surfaces is given by the inverse ratio of the exposures needed to photograph them, which from our experience is about 10^{-6}, corresponding to exposures of 100 sec at f 5.6 and 1/200 sec at f 32, respectively.

The scattering cross-section of particles in the size range inferred above is approximately equal to their geometrical cross-section πr^2, so that if n is their concentration we have

$$f(h) = hn\pi r^2 = 2 \times 10^{-6}.$$

The depth h of the layer occupied by noctilucent clouds is likely to be several km; for example, if the billows represent a convective process similar to that which produces the billows of tropospheric clouds, the depth of the layer involved is approximately $d/2.7$ (Scorer, 1951), where d is the separation of the billows, observed to be about 10 km. The billows, therefore, are likely to occur in a layer about 4 km deep, while the denser and much broader streaks and bands are probably even thicker. If we assume a value of about 6 km for h, then

$$nh\pi r^2 = 2 \times 10^{-6}, \text{ and}$$
$$n \approx 10^{-12}/r^2, \text{ or}$$
$$n \approx 10^{-2}/\text{cm}^3 \text{ if } r = 10^{-5}\ \text{cm},$$
$$n \approx 1/\text{cm}^3 \text{ if } r = 10^{-6}\ \text{cm}.$$

The corresponding space-densities of the material, assuming a particle density of 2.5, would be 10^{-16} and 10^{-17} g/cm³, respectively. These estimates can be regarded as representative of average conditions, and as unlikely to be in error by more than one power of ten. It would be very desirable to make better estimates by accurate quantitative study of particular clouds.

6. The substance of noctilucent clouds

Noctilucent clouds may be true ice clouds, or their particles may be composed of some other material.

In the ozone layer, and at higher levels up to about 55 km, the saturated vapour pressure

of water at the rather high temperatures encountered exceeds the air pressure, and a condensation of vapour is therefore impossible. At still higher levels, however, the temperature falls with increasing height, to reach a minimum at a kind of second tropopause at about 80 km, approximately the level of noctilucent clouds, and a condensation is possible there if the air is moist enough and the temperature sufficiently low. In the lower stratosphere of middle latitudes the air is practically always very dry, and water clouds are never observed apart from the exceptional wave clouds sometimes formed over the Norwegian and some other mountains in winter. The soundings of the British Meteorological Research Flight have shown that 3 to 5 km above the tropopause the frost-point tends to become constant at about $-115°$ F (193° A), corresponding to a relative humidity of less than 1 % (MURGATROYD AND OTHERS, 1955). If such air from these levels were raised or stirred up to the 80 km level, the mixing ratio, of order 10^{-3} g/kg, would be preserved, and the vapour density would become about 3×10^{-14} g/cm^3 (at 80 km the air density is about 3×10^{-8} g/cm^3 (WHIPPLE, 1952)). If this value is to exceed the saturated vapour density over ice, so that a condensation could occur, then the air temperature would need to be lower than about 145° A (accepting a value of the saturated vapour density extrapolated to this temperature from a formula regarded as thoroughly reliable in the range 173° to 273° A: O.M.I., 1951). Temperatures at the 80 km level have been deduced from instruments carried on rockets and from observations of meteors, and although they cannot be regarded as particularly accurate and were obtained mainly in latitudes lower than those in which the clouds are seen, they suggest a mean temperature of 220° to 230° A. WEXLER (1950) considers 170° A as a likely minimum value, and at this temperature the saturated vapour density is about 10^{-9} g/cm^3. It must therefore be said that a condensation in the 80 km level could be expected only if the temperature there became abnormally low, or if the air were abnormally humid, containing a large proportion of water vapour; nevertheless if the clouds there are regarded as very rare phenomena, the possibility of their representing a condensation cannot quite be excluded on these grounds.

A further circumstance makes it seem improbable. The rapid changes of the internal details of the clouds and the presence of billows suggest that the layer in which the clouds occur is well stirred and has a strong lapse-rate; since the layer is a few km deep, the temperature in its upper parts must be twenty degrees or so below that in the lower part. As the saturated vapour pressure decreases very rapidly with temperature, the space-density of the condensate in a cloud extending through such a layer must be nearly the saturated vapour density at the lower surface, that is, at the condensation temperature. Even if this temperature were as low as 145° A, the corresponding mean density of the ice cloud would be of the order of 10^{-14} g/cm^3, two or three orders of magnitude greater than the values deduced from the optical properties of the clouds.

Finally, the deduced concentrations of the cloud particles suggest a nucleated rather than a homogeneous condensation, so that if the clouds are regarded as ice clouds it is still necessary to consider upon what nuclei the particles formed. It is therefore natural to consider seriously other theories of the constitution of the clouds, which attribute them to concentrations of dust derived from the earth's surface or from outer space.

7. Noctilucent clouds and volcanic explosions

It has been suggested that the intensity and prevalence of noctilucent clouds in the years 1885 to 1894 are to be related to the eruption of Krakatoa in August 1883. There is no doubt that vast quantities of dust are introduced into the lower stratosphere by eruptive clouds. The final paroxysmal eruptions of Krakatoa occurred on the night of August 26, but already in the early afternoon the captain of the British ship *Medea*, over 75 miles away, could see the enormous cloud over the volcano, and from angular measurements estimated the height of its top as 17 mi, and later as 21 mi (34 km) (Report of the Krakatoa eruption, 1888). Since the wind at these levels carried the cloud away from the ship (towards the west), there is no reason to suppose that these are overestimates. Dust from the eruptive cloud spread across the earth and within three months was visible in western Europe as a tenuous layer even in full daylight, which produced spec-

tacular twilight phenomena. Estimates of the height of this dust cloud, based on the sun's depression when the red glow faded from the cloud, varied between about 15 and 30 km, the higher values being obtained mainly in the early history of the cloud, and the lower values after some months; amongst the observers was Jesse, who found mean heights of 12—15 km, later becoming 10 km.

Similar, though less dense clouds of dust were recently observed at such heights by aircraft over western Europe in July and August 1953 (JACOBS 1953), following an eruption in Alaska, and again in April and May of 1955, following an eruption in Kamchatka (BULL and JAMES, 1956). The appearance of the clouds and estimates of their height were also recorded by ground observers (LUDLAM, unpublished). It therefore seems likely that dust from volcanoes is injected mainly into the lower stratosphere, but it must be supposed that a sufficient proportion to account for the appearance of noctilucent clouds could become diffused up to much higher levels.

It is remarkable that the first observations of noctilucent clouds in the years 1885—1894 were not made until 2 years after the arrival of the volcanic cloud in Europe, and Vestine points out that the period 1880—1887 was one of strikingly frequent and brilliant displays of comets and meteors, and that the great eruption of Katmai, in 1912, was not followed by any increase in the frequency of reports of noctilucent clouds. On the other hand, SPANGENBERG (1949) has tried to associate more frequent reports in the years 1932—1935 with an eruption in the Cordilleras in 1932, which threw up a cloud to 25 km. However, major volcanic eruptions occur sufficiently often for some to lie within any chosen period, and the reports of noctilucent clouds are made so unsystematically and depend so much on favourable weather in a particular short season, that it is hardly surprising no close relation can be demonstrated between volcanic activity and the occurrence of the clouds.

8. Noctilucent clouds and meteor showers

On the morning of June 30, 1908, a great meteorite fell in Siberia; its mass has been estimated at some tens of thousands of tons, and the explosion it produced was recorded by barographs in England. During the same night an unusually brilliant display of noctilucent clouds was reported by many observers in Russia, Sweden, Denmark, Germany and England. G. A. Clarke, at Aberdeen, observed the sudden arrival of the tenuous clouds, accompanied by a marked brightening of the twilight, about two hours after sunset. From this observation WHIPPLE (1930) deduced the eastward speed of the cloud movement to have been about 85 m/sec, supposing the cloud to have originated in the trail of the meteorite. Apparently it was on this exceptional occasion that noctilucent clouds were reported (over Russia) from as far south as latitude 45° N (VESTINE, 1934); no references to the clouds or to unusual twilight brightness could be found in the meteorological logs in the U.S.A., and no such reports were made in southern Europe.

A witness of the Great Siberian Meteorite said that it left in the atmosphere 'a light bluish trail', and the smaller streaks left by meteors large enough to be seen in daylight also have the pale silvery colour which is characteristic of noctilucent clouds; evidently, therefore, the particles in meteor trails have about the size of those composing the clouds. This conclusion is supported by the work of FESSENKOV (1949), who has discussed the properties of the trail left by another large meteorite, of estimated mass a few hundred tons, which also fell in Siberia, on 12 February 1947. Broken material was found in small craters, and with the help of mine detectors many small pieces, including a multitude of spheres of radii 15—20 μ produced by thin sprays of solidifying iron, were found in the soil. At first the trail left in the atmosphere was about 1.5 km in diameter, and was so dense as to obscure the sun, but soon afterwards, as it diffused, the sun shone through as a red disc. The particles were therefore noticeably more transparent to red than to blue light. From these observations Fessenkov deduces that the diameter of the particles must have been less than 10^{-5} cm, and that their concentration in the original trail was about $7 \times 10^5/\text{cm}^3$; since the total volume of the trail was originally about 70 km^3, it contained nearly 200 tons of iron. If the particles had diameters considerably smaller than 10^{-5} cm, these figures for the space density and total mass would be substantially increased. Since if noctilucent clouds are supposed to

contain particles of diameter 10^{-5} cm their concentration would be about $10^{-2}/cm^3$, such a trail could diffuse horizontally to a diameter of 7,000 km before the mean particle concentration fell below that inferred for noctilucent clouds. Although the denser parts of a large meteorite trail lie below the 80 km level, clearly such trails could form a significant source of the substance of noctilucent clouds, divided into particles of the appropriate size. The particles are presumably produced by condensation of material boiling away from the incandescent meteorite, and perhaps also by the crumbling of its substance into the component cosmic particles. The much larger spherules sometimes found are drips from the molten but unvaporized parts of meteorites which reach the ground, and characteristically have radii of some tens to hundreds of microns.

Meteorites of the great size of those discussed above are very infrequent. Recently Bowen (1953) has claimed that there is a very close relation between the occurrence of noctilucent clouds and recurrent meteor showers. Bowen compared the dates of the prominent annual meteor showers of June and July, according to Lovell and Clegg (1952), with the dates of maximum frequency of reports of noctilucent clouds, according to the diagram of Vestine (1934) (Fig. 4). This diagram contains two curves, one of which shows, for reports collected by Vestine from the period 1885—1933, the total number of years in which reports have been made on particular dates, and the other of which shows the total number of reports on particular dates. The latter curve was alone considered by Bowen; it has several peaks which in seven cases out of eight correspond to or precede by 1 or 2 days the dates quoted for the prominent meteor showers. Some authors have since supposed that in this way Bowen has established a relation between the two phenomena. However, Whipple and Hawkins (1956) state that the dates of the principal meteor streams used by Bowen are inaccurate, and that when more modern data are used the only notable coincidence with the dates of the maxima on Vestine's curve occurs on June 30, corresponding to a weak meteor stream. On this date Vestine's curve indicates 17 reports, which an examination of his lists shows to contain probably 12 from the occasion of the fall of the Great Siberian Meteorite on June 30, 1908. Thus there is good reason to discount even this coincidence. Nor is the correspondence improved by considering Vestine's curve which shows the number of years, rather than the number of observers, when the clouds have been recorded, which might be regarded as the more appropriate curve to discuss. On this curve 5 occasions, and one the other curve 7 reports, would be sufficient to produce a peak, so that the significance of those shown in the diagram must be regarded as very doubtful. Moreover, it must be realized that the observations which it summarizes were made over a large range of latitudes, within which the viewing conditions vary considerably from place to place on the same date. For example, even excluding the peak of June 30, Vestine's diagram shows that the observations are notably most frequent about the end of June; however, if a sufficient number of observations from high latitudes were included, this maximum would disappear, for near midsummer the clouds there are invisible, even if present. Evidently only an extended series of observations in one locality or latitude, including reports of occasions when no clouds could be seen in apparently favourable viewing conditions, can provide a satisfactory basis for relating the occurrence of the clouds to any other phenomena.

The observations made in Sweden (Fig. 12) suggest that within the best viewing season there are clear nights when the clouds are definitely not observed, separating the occasions when they are seen. The occurrences are not obviously related to any meteor streams; the principal stream at this season is that of the Perseids, which begin to arrive about August 6 and reach a maximum between about August 10 and 14. This meteor shower, in common with the others, is spread over a few days, and cannot be associated with a particular date (Lovell, 1954). On the basis of the available evidence, therefore, it cannot be concluded that there is any clear relation between the occurrence of noctilucent clouds and the arrival of meteor showers.

9. The accretion of interplanetary particles by the earth

The rate of arrival of visible meteors during the major showers rises by only a factor of

about 2 above that of the sporadic meteors (Lovell, 1954), so that the showers can be neglected in comparison with the sporadic meteors as a possible source of the material of noctilucent clouds.

Visible meteors are produced by particles of mainly stony or iron-nickel composition, and of radius exceeding about 100 μ, which enter the atmosphere at speeds of some tens of km/sec and become incandescent. Particles smaller than this are also fused, but emit too little light to be visible to the unaided eye, while particles of radius less than a few microns are decelerated in the atmosphere without reaching melting-temperatures (1 200—1 700°A) (Whipple, 1950, 1951). The larger meteorites leave trails which often are visible for some minutes or even longer as a dark smoke, or, under favourable lighting, as pale streaks closely resembling the noctilucent clouds. It seems, also, that many meteors, particularly the fainter ones, have a low density (0.2 to 1.0 g/cm^3), and disintegrate under aerodynamic pressure and heating by a progressive crumbling (Jacchia, 1955), so that they are regarded as porous and fragile structures, probably consisting of mineral fragments embedded in ices of such substances as H_2O, CO_2 and NH_3. However, particles of the size (10^{-5} to 10^{-6} cm) inferred to predominate in noctilucent clouds must be supposed to enter the atmosphere either directly with this size, or to be produced in the atmosphere by condensation of vapourized material in the wakes of meteorites having dimensions which exceed several microns. The visible paths of most meteors lie between heights of about 100 and 50 km above the ground, so that their condensation products are distributed over a layer extending somewhat above and below the level of noctilucent clouds.

The fall-speeds of small particles at these levels have been estimated by Link (1950), using Stokes Law with Millikan's correction; he obtained values proportional to the particle radius r and density δ, and inversely proportional to the air density ϱ. At the 80 km level he found a fall speed of a few cm/sec for particles of radius 10^{-5} cm. On the other hand, the mean free path of the air molecules at this height is about 4 mm, and therefore so much greater than the particle sizes that it seems preferable to use Knudsen's (1934, p. 33) empirical data for the resistance K of rarified air to the motion of a sphere. He measured K in air at pressures extending down to 0.14 dyne/cm^2, and expressed his results by a formula

$$K = 6\pi\eta r v(1 + 0.683\varkappa + 0.354\varkappa/e^{1.845\varkappa})^{-1},$$

where η is the coefficient of viscosity and $\varkappa = \lambda/r$, λ being the mean free path obtained from the expression

$$\lambda = \sqrt{(\pi/8)}\eta/0.30967p\sqrt{\varrho_1},$$

where p is the pressure and ϱ_1 is the density at 1 dyne/cm^2.

If $\varkappa \ll 1$ the formula reduces approximately to

$K = 6\pi\eta r v\varkappa$, and hence

$v = \varkappa v_s$,

where v_s is the Stokes Law fall-speed.

At the 80 km level we may assume $T = 200°$A, $\eta = 1.3 \times 10^{-4}$ g/cm sec, and $p = 17$ dynes/cm^2; with $\varrho_1 = 1.74 \times 10^{-9}$ g/cm^3 we then have $\lambda = 0.37$, and for particles of density 2.5 g/cm^3

$$v = 4.2 \times 10^6 r^2 \varkappa.$$

If $r = 10^{-5}$ cm, $\varkappa = 3.7 \times 10^4$ and $v = 16$ cm/sec.
If $r = 10^{-6}$ cm, $\varkappa = 3.7 \times 10^5$ and $v = 1.6$ cm/sec.

These estimates differ by less than 50% from values deduced from ordinary kinetic theory of gases (Witt, private communication).

These values incidentally lend some support to the upper size limit deduced for the particles of noctilucent clouds on optical grounds, for it seems inconceivable that the clouds could possess persistent identifiable details if the fall speeds of their particles reached values in m/sec.

Given these fall speeds and the inferred space-density of noctilucent clouds, we can compute the density in space of meteoritic material sufficient to provide the source of the clouds and compare this with the density believed to exist, since meteoritic material arrives from interplanetary space with speeds within the range of about 11 km/sec (the speed of fall due to the earth's gravitation) and about 70 m/sec (compounded of the earth's orbital speed about the sun and the speed of escape from the sun at the distance of the earth).

From the observed properties of visible meteors, the size distribution of the particles over the range of radii from about 50 μ to 0.5 cm can be inferred. On the basis of this size distribution law, extrapolated to include both smaller and larger meteorites, Lovell (1954)

estimates the annual accretion of meteoritic material by the earth to amount to about 500 tons, corresponding to an average density in interplanetary space of between 10^{-25} and 10^{-24} g/cm^3.

If material in the higher concentration were to enter the earth's atmosphere at 32 km/sec, and were wholly converted into particles of radius 10^{-6} or 10^{-5} cm, settling at their terminal speeds, its density at the 80 km level would become 2.10^{-18} or 2.10^{-19} g/cm^3, respectively. If these values are compared with those deduced for noctilucent clouds composed of such particles (about 10^{-17} and 10^{-16} g/cm^3, respectively), we see that they are smaller by nearly one to three orders of magnitude. Although the magnitudes of the compared quantities are each liable to be in error by as much as one power of ten, the values obtained for the density of the meteoritic material are likely to be too high rather than too low. The available evidence therefore cannot be considered with any confidence to present ordinary meteorites as a sufficient source of the material of noctilucent clouds. Other possible sources must also be considered.

10. The zodical light cloud

There is some evidence that in the meteor showers the numbers of very small particles are even less than indicated by extrapolation of the size distribution law amongst the visible meteors. One probable reason is that the small particles in meteorite swarms become separated from the larger. On the other hand there is also reason to suppose that the accretion of sporadic micrometeorites, too small to produce visible meteors, is very much greater than this law would suggest. This evidence comes mainly from the study of the sun's outer corona and the zodiacal light.

The zodiacal light is a wedge of diffuse, feeble light pointing upwards from the horizon along the position of the ecliptic during the dark hours, when the depression of the sun exceeds about 17°; in intensity it is comparable with the light of the Milky Way (see, e.g., MINNAERT, 1940). It can be seen easily on clear, moonless nights in low latitudes, where it is always steeply inclined to the horizon, but it is more difficult to observe in middle and high latitudes.

The intensity of the light increases towards the sun, and VAN DE HULST (1947) suggested that this increase extends continuously into the outer corona of the sun. The light of the corona is composed of two parts, a *K*-component which is highly polarized and contains no spectral lines, and an *F*-component which has little or no polarization and contains Fraunhofer lines. The *K*-component decreases very rapidly in intensity away from the sun, but the *F*-component intensity decreases more slowly, and dominates the outer corona. The zodiacal light also contains the Fraunhofer lines, and both it and the *F*-component of the corona are attributed to scattering and diffraction of sunlight by a cloud of interplanetary particles disposed as a rather shallow disc lying in the plane of the ecliptic. At the time of van de Hulst's suggestion the only measurements available were those of the intensity of the zodiacal light in the range of solar elongations from 40 to 180°, and those of the intensity of the F-component of the corona, out to about 1° 20′ from the sun. Recently, however, RENSE and others (1953) have measured intensities in the range of solar elongations from 5 to 13° (from an aircraft flying at 10 km during a total solar eclipse); their results confirm van de Hulst's suggestion, forming a smooth link between the measurements of the corona and the zodiacal light.

ALLEN (1946) showed that the intensity of the outer corona and the zodiacal light, and also the distribution of intensity with solar distance and wavelength, could be explained by the diffraction and scattering of sunlight from particles of radius about 10^{-3} cm, whose concentration, assumed inversely proportional to the distance from the sun, is about 3×10^{-15}/cm^3 at the distance of the earth, representing a space density of about 6×10^{-23} g/cm^3. His measurements in the outer corona showed no indication of Rayleigh scattering such as would be produced by much smaller particles; on the contrary, there was some evidence of a reddening of the *F*-component near the sun. Recently BLACKWELL (1952) found a considerably greater increase towards the sun in the infra-red than in the violet, indicating the presence of particles of radius larger than about 2×10^{-3} cm. Van de Hulst assumed a size distribution of the scattering particles similar to that of meteoritic material, and a uniform

concentration of the particles, considering that the increase of intensity towards the sun is due to the form of the scattering function, rather than to changes in the particle concentration. Since part of his scattering and diffracting material is distributed in larger, less efficient particles, he deduces a space-density, of 5×10^{-21} g/cm^3, which is considerably greater than the value of about 6×10^{-23} g/cm^3 corresponding to Allen's result. If van de Hulst's figure is accepted as probably more realistic and the thickness of the dust cloud perpendicular to the plane of the ecliptic is about 0.1 AU, the total mass of the particles within the earth's orbit becomes 5×10^{12} tons, or about 10^{-9} times the mass of the earth.

It is interesting to consider the origin and fate of this material. WHIPPLE (1955) considers four possible sources, of which only asteroidal collisions and cometary disintegrations appear significant. Of the particles which are produced in these ways those of radius less than about 2×10^{-5} cm are expelled from the solar system by the radiation pressure of sunlight. Larger particles spiral towards the sun under the Poynting-Robertson effect (see, for example, WYATT and WHIPPLE, 1950), at a cosmically rapid rate. Whipple deduces that to balance this loss and to maintain the outer corona and zodiacal light at their present brightness a source of about 1 ton/sec of meteoritic material is required, and quotes Fessenkov as having previously arrived at a similar conclusion by a less general theory. This figure may be compared with PIOTROWSKI's (1953) estimate of between 20 and 600 tons/sec of pulverized material which is produced by collisions between the bodies in the asteroidal belt. Probably, however, the bulk of this material is in the form of rather large fragments. According to WHIPPLE (1954) most of the small particles producing photographic meteors are derived from cometary debris (according to Whipple's model, a comet nucleus is composed of solid particles embedded in "ices" of substances such as H_2O, NH_3 and CO_2, which are partially vapourized on approach to the sun, so that the comet progressively disintegrates during successive orbits). Whipple estimates that particles in the debris of size greater than 10^{-2} cm are exposed to a high probability of destructive collisions (particles of size larger than this are also unlikely to escape accretion with Jupiter during their passage of its orbit, according to ÖPIK, 1951), but that the cometary source of material for the zodiacal light cloud could readily amount to about 3 tons/sec, which is sufficient to sustain the cloud.

The available astronomical evidence, though rather tenuous, is thus consistent with the presence of an interplanetary dust cloud, in which the particles have radii mainly in the range 10^{-3} to 10^{-2} cm, and whose concentration corresponds to a space density of about 5×10^{-21} g/cm^3. Such particles entering the atmosphere would be expected to produce very faint telescopic meteors in numbers about 10^4 greater than would be deduced from extrapolation of the size distribution law for meteors visible to the unaided eye. So far this great increase has not been demonstrated, but indirect evidence does suggest that the previously accepted accretion rates of meteoritic material are too small. For example, the impacts of small particles (of size a few microns) upon high-altitude rockets are greater than was anticipated (see WHIPPLE, 1952, p. 22), and so also is the abundance of iron and nickel in deep-sea sediments (PETTERSON and ROTSCHI, 1950; the significance of these deposits has been questioned, however: see, for example, ÖPIK, 1955), and the rate at which magnetic material of presumably meteoritic origin falls to the ground (THOMSEN, 1953; here also, however, the origin of the material is not certain: SCHAEFER, 1955, attributes most of it to industry).

If the estimate of the composition of the zodiacal light cloud is accepted, and the particles are supposed to evaporate in the high atmosphere and subsequently condense into particles of radius 10^{-6} or 10^{-5} cm, the space densities of these particles at the 80 km level estimated on the previous reasoning would be as follows:

for $r = 10^{-6}$ cm, space density $= 10^{-14}$ g/cm^3;
for $r = 10^{-5}$ cm, space density $= 10^{-15}$ g/cm^3.

Both these values exceed the space densities inferred for noctilucent clouds (10^{-17} and 10^{-16} g/cm^3, respectively). Since we have used van de Hulst's rather than Allen's estimate of the density of the zodiacal light cloud, and since a proportion of its particles would be small enough to escape boiling in the atmosphere, the values are likely to be too high rather than too small. Nevertheless it may be concluded

that interplanetary dust is a possible source of the material of noctilucent clouds. It also seems possible that if this dust contained a substantial proportion of iron, or if it were charged by solar radiation, it might enter the earth's atmosphere preferentially in high latitudes, thereby locally increasing in concentration.

11. Dust in the high atmosphere

Other evidence is available to suggest the presence of considerable quantities of dust in the stratosphere. For example, observations of the earth's shadow during lunar eclipses suggest that the atmosphere contains obscuring matter to levels well within the stratosphere, which LINK (1950) has proposed to be meteoritic dust extending from a height of 50 to 100 km down to the ground. Photographs of Mars and of Venus give ultra-violet images greater, by 3—6% and 2% respectively, than the infra-red images. According to LINK (1950), Fessenkov and Menzel have proved that an explanation in terms of molecular diffusion is not possible, and the discrepancies are tentatively explained by the presence of similar meteoritic dust layers. The smaller gravitational constants on these planets lead to a level of meteorite fusing which is farther above the ground than on earth, favouring a deeper dust layer. On the other hand on Jupiter, for example, the layer could be only some tens of km deep and could not produce noticeable effects. BOUSKA and SVESTKA (1959) have collected measurements of the size of the earth's shadow on the moon during lunar eclipses; the mean from 33 observations shows the shadow to be larger than the earth by about 2%, individual values varying from 1.7 to 2.5%. Bouska and Svestka were unable to find any significant relation between the enlargement and the declination of the sun or moon, the parallax of the sun, or the solar activity, but do claim a correlation with the occurrence of meteor showers.

From an analysis of observations of solar intensity at Mount Wilson in the period 1908—1920, ZACHAROV (1952) deduces an annual increase in atmospheric absorption by a few per cent over about four weeks including the latter half of August, which he associates with the arrival of the Perseids in

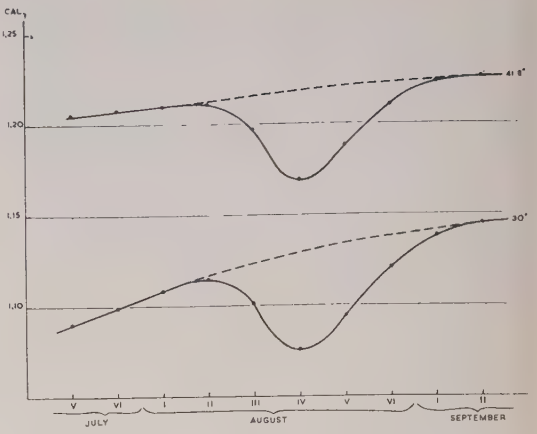


Fig. 13. Average solar intensity for each 5 days for a sun's elevation of 30° (lower curve) and 41.8° (upper curve), from 3282 and 2239 observations made at Yakutsk (1931—42), Pavlovsk (1913—37), Kursk (1928—35), Rostov (1930—37), Tbilissi (1928—37), Tashkent (1926—37) and Samarkand (1931—37). After Kalitin (1944).

mid-August. KALITIN (1944) showed that a similar decrease in the solar intensity at the same time of year is evident in the actinometric measurements of a number of widely-separated stations in Russia (Fig. 13); the coincidence is extraordinary. Presuming the effect is produced by the entry of a dust cloud into the atmosphere LINK (1955) infers the particle radius to be about 10^{-5} cm, and that the dust arrives at about the rate deduced above for accretion with the zodiacal light cloud. However, since the optical density of such a cloud would be several orders of magnitude greater than that of noctilucent clouds, it seems remarkable that its annual arrival is not a visible and celebrated event. Dust clouds from the catastrophic eruptions of Krakatoa (1883), M. Pelée (1902) and Katmai (1912), produced more prolonged decreases in solar intensity of about the same amount, but equally produced twilight phenomena of abnormal intensity (GRUNER and KLEINERT, [1927). Among these phenomena the purple light, which is attributed to particles, of dimensions a few microns, in the lower stratosphere, is especially enhanced in such disturbed periods. It may be significant that Fig. 14, taken from GRUNER and KLEINERT (1927), shows that abnormally intense purple lights are notably more frequent in August (when, however, convective lifting of dust from the ground is intense) than at other times

of the year. Although, therefore, the evidence of Zacharov and Kalitin may need some other interpretation, their suggestions deserve further investigation. It may be noted that the recovery of the actinometer measurements after four weeks would suggest, if the disturbance is due to the arrival of a dust cloud, that the dust is substantially precipitated from the atmosphere within this period.

Direct confirmation of the presence of dust concentrations in the lower stratosphere comes from aircraft observations (e.g., PACKER and LOCK, 1950), and the photometric twilight observations of BIGG (1956) provide a strong indication of the persistent presence of concentrations of dust below the 80 km inversion. PENNDORF (1954) has assembled observations of particle concentration at various levels in the troposphere which show that the strong decrease with height ceases above about 4 km, and that thereafter the ratio between the concentration and the air density (a ratio we might expect to be preserved during the stirring of air to different levels) tends to increase with height, suggesting that at least a proportion of the particles in the upper troposphere have a source at a higher level. This result receives some support from the particle concentrations observed to occur in water clouds at the cirrus levels and in the spectacular mother-of-pearl clouds which occasionally form some 25 km above the Norwegian mountains. From coronae seen on freshly-forming cirrus I deduced that the concentration of ice crystals in the clouds is about 7/cm³ (LUDLAM, 1956). Since these clouds were formed in moderately strong updraughts this probably also represents practically the total concentration of condensation nuclei at these levels, in accord with the values inferred by Penndorf from flight measurements of solar radiation. Thus at the cirrus levels (about 10 km) the ratio $C = n/\varrho$, where n is the concentration of small particles, has a value of about $7/5 \times 10^{-4}$/g, or 1.4×10^{4}/g. An observation by Størmer (1948) of the angular radius (15°) of the first red ring of a lunar corona in mother-of-pearl clouds, indicates spherical cloud particles of diameter about 2.5 μ. At the temperature of these clouds (about −80° C) the concentration of water condensed during a rise of a few hundred metres above the condensation level amounts to about 10^{-10} g/cm³; if this is divided uniformly (as justified by the purity of the diffraction colours) amongst droplets of the appropriate radius, their concentration becomes about 1/cm³. Since these wave clouds are also produced in updraughts of at least moderate strength, we can equally assume this value to represent the total concentration of condensation nuclei. At 25 km the air density is approximately 6×10^{-5} g/cm³, so that here the ratio C becomes 1.6×10^{4}. At the level of the noctilucent clouds the air density is about 3×10^{-8} g/cm³, and the inferred concentrations of particles vary from 10^{-2}/cm³ (for a radius 10^{-5} cm) to 1/cm³ (radius 10^{-6} cm), corresponding to values of C of 3×10^{5} to 3×10^{7}. These values are rather higher than those deduced for the 10 km and 25 km levels, but the latter are so nearly the same and sufficiently close to the lower estimate for the 80 km level to encourage the view that the distribution of dust throughout the stratosphere is consistent with the existence of efficient mixing movements, and not inconsistent with a meteoritic source.

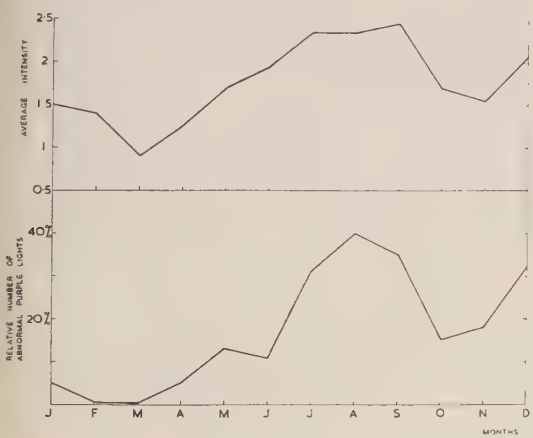


Fig. 14. Average intensity of purple light (on scale from 0 to 5), and relative number of abnormally strong purple lights, Berne, 1905—1912 (after Gruner and Kleinert, 1927).

The occurrence of stirring motions in the lower stratosphere, where the air is statically very stable, and even in the upper stratosphere, where the air is stable for dry adiabatic ascent and condensation is impossible, is consistent with the uniform composition of the atmosphere but is difficult to reconcile with the usual thermodynamical arguments. RAKIPOVA (1947), however, has pointed out that dust

clouds may be agencies for producing upward movement of air, by intercepting sufficient sunshine to become appreciably warmed. Air containing a dust cloud could thus commence to rise through a clearer environment, and, since it could carry its dust with it, may remain warmer and more buoyant than the surroundings. In these circumstances vertical movements are possible in statically stable, even isothermal atmospheres. Rakipova calculates the concentrations of dust particles required to produce significant effects and finds, for a radius 10^{-5} cm, a concentration of 1/cm^3 at 80 km and about 10/cm^3 at 30 km. These estimates may be obtained in the following way. At the 80 km level, for example, where the air density is 3×10^{-8} g/cm^3, solar energy is intercepted by particles (assumed to have a low albedo) at the rate of about $N\pi r^2$ cal/cm^3 min, where N is the concentration of particles of radius r, or about 200 Nr^2 cal/cm^3 hr. If this energy is communicated to the surrounding air it is sufficient to warm it at the rate of about $10^3\ Nr^2/\varrho$ °C/hr, or about 2×10^{10} Nr^2 °C/hr at the 80 km level, so that for $r = 10^{-5}$ cm and $N = 1$/cm^3 the rate of heating becomes appreciable and equivalent to a temperature rise of 2° C/hr. In a lapse-rate of 6° C/km this could correspond to a vertical movement of a km in 3 hours, at a speed of about 1/3 m/sec. Evidently for this process to be important the concentration of particles would need to be at least two orders of magnitude greater than that deduced to occur in noctilucent clouds (10^{-2}/cm^3 for $r = 10^{-5}$ cm). Since meteoritic material is unlikely to enter the atmosphere in a uniform stream, such high concentrations may sometimes occur in clouds which, if sunlit for a sufficient period, could by this means rise several km to the 80 km level. Above the base of the inversion there, the ascending speeds would be substantially reduced, and insufficient to lift the particles of the cloud, whose fall speeds amount to at least several cm/sec, so that the ascent would cease.

Particles which settle from the stratosphere into the troposphere aggregate with the condensation nuclei derived from the earth's surface and become involved in precipitation processes. The observations quoted above suggest that in the upper troposphere, where the total concentration of nuclei is about 10/cm^3, solid meteoritic particles may often outnumber the nuclei of terrestrial origin. In the lower troposphere, where their concentration could not be significantly greater, they could comprise only an insignificant fraction of the condensation nuclei, although, in view of their silicate components, they might conceivably be an important source of freezing nuclei in clean air masses (on the other hand SCHAEFER, 1955, found that particles of magnetic dust collected in the open air, and particles from the smoke of a fused *metallic* meteorite, were not particularly effective ice nuclei).

12. The formation of noctilucent clouds

Some of the observations made in Sweden, particularly when noctilucent clouds developed after midnight, strongly suggested that the clouds formed within periods of several minutes in regions which previously, although apparently favourably illuminated, contained no discernible details. GRISHIN (1955) has remarked that the featureless "flare" can often be detected before the appearance of definite clouds, and thus also implies that clouds may form in the field of view, not necessarily simply arriving from places in which they already existed, or being revealed by improving viewing conditions.

The readiest interpretation of the formation in a restricted layer of clouds having definite details is in terms of a condensation process in air which is ascending or otherwise being chilled. However, we have already discussed difficulties in the way of this explanation. To reiterate, a condensation can occur only if the air were much moister or much colder than we have reason to expect, and even then the brightness of the clouds would suggest the condensation to be only partial, while the problem of accounting for the condensation nuclei would remain. If, moreover, the clouds were to appear by the growth of their particles, their first details might be expected to have a distinct blue colour which becomes steadily more white. Such a change of colour has only once been reported (PATON, 1954) and cannot be regarded as characteristic. Equally, then, it seems unlikely that the clouds become visible by a condensation (at vapour pressures below the saturation value) upon hygroscopic components of the nuclei.

If condensation of water vapour plays no part in the formation of the clouds, it must be considered how dust, probably present throughout the stratosphere, can become so disposed to produce distinct clouds which are always measured to be at a height of about 80 km. It seems likely that the formation of a *haze top* at the level of the inversion there is essential.

Pronounced haze tops are common in the troposphere, and are usually associated with inversions; in a layer of steep lapse-rates beneath the inversion particles of dust or other nuclei are stirred up from their sources near the ground, but above the base of the inversion stirring movements are suppressed by the strong static stability, and the concentrations of the particles are suddenly reduced by a factor of 10 to 100. Moreover, beneath the inversion the relative humidity increases with height, so that the size of the particles, which are all more or less hygroscopic, reaches a maximum at the haze top; immediately above, the air is usually very dry and the particles there are inappreciably swollen by condensation, so that the presence and distribution of water vapour are important in sharply defining the top of the haze layer. When wave motions disturb the layer, concentrations of haze become visible from below to an observer who looks along low angular elevations, for then adjacent lines of sight pass through layers of appreciably different optical thickness (Fig. 15). Patterns of dark haze billows are seen against a twilight background, or, if the haze top is still illuminated, bright billows are seen separated by darker spaces. Such billows are a feature of the dust cloud in the lower stratosphere, which cannot be seen overhead, but only at rather low elevations.

It is probable that the details of noctilucent clouds are produced in a similar way by disturbances in a haze top at the 80 km inversion.

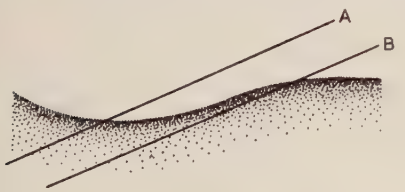


Fig. 15. A haze layer with a sharply-defined and undulating top presents visible features at low elevations, for there are marked differences in optical thickness along adjacent lines of sight, such as A, B.

It would not be surprising if all the sharply-defined clouds, which are naturally those selected for measurement, were necessarily associated with the haze top, and that other clouds are present which are considerably lower, but have more diffuse edges, and whose height is therefore never determined. The distinct "shadows" which sometimes seem to be cast by the denser cloud features (Fig. 1) almost certainly represent irregularities in a haze top, for the optical thickness of the clouds is so small that they could not possibly cast real shadows.

The production of a haze top at 80 km seems to require the ascent of dust clouds from below. Wexler (1950) has discussed observational evidence of very steep, even superadiabatic lapse rates, in the layer between about 55 and 80 km. These arise during the daytime by solar heating of the ozone layer; although the ozone is concentrated at lower levels, the diurnal temperature change has a maximum at about 50 km (Johnson, 1953). It must be considered that the noctilucent clouds are observed only at a season and in latitudes such that a very prolonged heating of the ozone layer has occurred: at mid-summer in high latitudes the sun does not set on the top of the ozone layer. At this season, therefore, convective movements are particularly likely, and we may suppose that the ascending motions will occur preferentially in dust clouds. Consequently a concentration of dust may arise beneath the 80 km inversion, where the dusty air loses its buoyancy and spreads horizontally. If the dust has entered the stratosphere from below, as in eruptive clouds, a pronounced haze top is readily produced. If it has arrived from interplanetary space as the condensation product of fused meteors, it is not possible to assess the change of concentration which occurs across the 80 km inversion without information on the accumulation which occurs in the convective layer and on the levels at which the condensation takes place. Bright meteors produced by particles of mass several grams arriving at about 40 km/sec are first visible at a height of about 100 km and disappear at about 70 km; larger particles are able to penetrate to lower levels, and so considerable proportions of their condensation products are introduced below the 80 km level. This circumstance would help in the producuon of a well-defined

haze top; on the other hand, estimates of the evaporation of particles of the small size inferred to compose the zodiacal light cloud, based upon the considerations summarized by WHIPPLE (1943), indicate that they must be evaporated entirely above the 80 km level. It therefore seems likely that only meteorological processes could account for the formation of a haze top or a marked concentration of small particles near the 80 km level.

It is interesting that dust clouds, which in sunlight may cause an appreciable warming of air, may equally be responsible for a significant cooling of air when they lie in the earth's shadow. Since the long-wave radiation emitted by the earth and its atmosphere has an equivalent black-body temperature of about 245° A, a particle in the high atmosphere (which radiates in all directions but receives radiation from the hemisphere below) is in radiative equilibrium at a temperature of about $245/2^{1/4}$, or 206° A. The air temperature in the upper stratosphere, at least to within a few km of the 80 km inversion, exceeds this value, so that dust clouds may behave as a cooling agent and promote sinking motions. Conceivably, in the high latitude summer, when the earth's shadow moves mainly latitudinally, there may be a tendency for a circulation in the vicinity of the shadow, in the layer from 50 to 80 km, as a result of a predominance of ascending motions in a belt a few hundred km broad in the sunlit region, and of descending motions in a similar belt inside the earth's shadow. Such a circulation would involve a component of the motion at the 80 km level directed away from the sun, as seems to be observed in the motions of noctilucent clouds.

13. Conclusion

It appears very unlikely that the condensation of water vapour plays any part in the formation of noctilucent clouds. They are probably composed of small solid particles, which may enter the stratosphere from below during major volcanic eruptions, or which may be produced in the stratosphere by the condensation of gases in the wakes of meteors. During the high latitude summer steep lapse rates develop in the upper stratosphere because of the prolonged heating of the ozone layer, and convective motions, possibly aided by solar warming of particle clouds, distribute the particles throughout a layer extending up to the 80 km inversion. Irregularities and wave disturbances in the top of the dusty layer produce visible clouds. The clouds may appear predominantly in regions near the earth's shadow at midnight, when adjacent belts of the atmosphere remain for some hours on either side of the shadow, so that a circulation is established in the upper stratosphere in which the sunlit air rises and the shadowed air sinks. This circulation and the haze top at 80 km which favour the appearance of the clouds may not develop in other latitudes and seasons because of the absence of steep lapse rates and because of the rapidity with which the earth's shadow moves through the atmosphere. The tendency for clouds to appear after rather than before midnight may be related to the diurnal variation in the accretion of interplanetary particles, but this and some other cloud properties need more thorough investigation. Noctilucent clouds may interest astronomers as evidence of the accretion of interplanetary material, and are important to meteorologists as indicators of atmospheric processes affecting the stratosphere, and as reminders that a considerable proportion of the condensation and freezing nuclei of the upper troposphere may arrive from outer space.

In future observations of noctilucent clouds it will be important to record occasions when they are absent in apparently favourable viewing conditions. It would be most desirable to establish a network of observing stations, to examine the true areal extent of the clouds and the simultaneity or otherwise of their appearance in different parts of the world. The ethereal beauty of the clouds on the rare occasions of good displays is a recompense for many hours of attentive but fruitless observing.

Acknowledgments

The observations of noctilucent clouds were made as part of a research programme in cloud physics under development at the International Institute of Meteorology in Stockholm. This programme was supported by Statens Tekniska Forskningsråd, Sweden, and the Munitalp Foundation, U.S.A. Grateful acknowledgment is made of the assistance of staff of the

Institute, and thanks are due to Mr. J. Paton, for making available a summary of his observations, and to Dr. R. M. Goody for illuminating comments and discussions. G. Witt gave helpful advice on the calculation of fall speeds of particles in rarified gases.

REFERENCES

Allen, C. W., 1946: The spectrum of the corona at the eclipse of 1940. *Mon. Not. R.A.S.*, **106**, p. 137.

Arago, F., 1854: *Complete works*. **4**, p. 73.

Archenhold, F. S., 1928: *Die leuchtenden Nachtwolken...*, Das Weltall. 27. Jahrgang.

Blackwell, D. E., 1952: A comparison of the intensities of infra-red and violet radiation from the solar corona ... *Mon. Not. R.A.S.*, **112**, p. 652.

Bigg, E. K., 1956: The detection of atmospheric dust and temperature inversions by twilight scattering. *J. Met.*, **13**, p. 262.

Bouska, J., and Svestka, Z., 1950: On the variation of the enlargement of the earth's shadow during lunar eclipses, *Bull. Astr. Inst. Czecho.*, **2**, p. 6.

Bowen, E. G., 1953: The influence of meteoritic dust on rainfall. *Austral. J. Phys.*, **6**, p. 490.

Bull, G. A., and James, D. G., 1956: Dust in the stratosphere over western Britain on April 3 and 4, 1956. *Met. Mag.*, **85**, p. 293.

Fessenkov, V. G., 1949: The mass of the atmospheric residue of the Sikhote-Alin meteorite. *Dok. Akad. Nauk SSSR*, **66** (translation T 133 R, Defence Sci. Inf. Service, Canada, 1954).

Grishin, N. I., 1955: On the structure of noctilucent clouds. *Met. i Gidr.*, No. 1, p. 23 (in Russian).

Gruner, P., and Kleinert, H., 1927: *Die Dämmerungserscheinungen*, Probleme der Kosmischen Physik, Henri Grand, Hamburg.

Jacchia, L. G., 1955: Physical theory of meteors. *A. J.*, **121**, p. 521.

Jacobs, L., 1954: Dust cloud in the stratosphere. *Met. Mag.*, **83**, p. 115.

Jesse, O., 1896: Die Höhe der leuchtenden Nachtwolken. *Astr. Nachr.*, **140**, p. 161.

Johnson, F. S., 1953: High-altitude diurnal temperature changes due to ozone absorption. *Bull. Amer. Met. Soc.*, **34**, p. 106.

Kalitin, N. N., 1944: Cosmic dust according to actinometric measurements. *Comptes Rendus (Doklady) Acad. Sci. URSS*, **45**, p. 375.

Khvostikov, I. A., 1952: Silvery clouds. *Priroda*, Moscow, **5**, p. 49 (in Russian).

Knudsen, M., 1934: The Kinetic Theory of Gases. Methuen, London.

Krakatoa Committee, 1888: *The eruption of Krakatoa and subsequent phenomena*. Trübner and Co., London.

Link, F., 1950: Couche de poussières météoriques dans une atmosphère planétaire. *Bull. Astr. Inst. Czecho.*, **2**, p. 1.

— 1955: Contribution of meteoritic material to atmospheric absorption, "Meteors", Special Suppt. *J. Atm. Terr. Phys.*, **2**, p. 36.

Lovell, A. C. B., 1954: *Meteor Astronomy*, Clarendon Press, Oxford.

— and Clegg, J. A., 1952: *Radio Astronomy*, Chapman and Hall, London.

Ludlam, F. H., 1956: The forms of ice clouds, II. *Quart. J. R. Met. Soc.*, **82**, p. 257.

Minnaert, M., 1940: *Light and colour in the open air*, G. Bell, London.

Murgatroyd, R. J., Goldsmith, P., and Hollings, W. E. H., 1955: Some recent measurements of humidity from aircraft up to heights of about 50,000 ft over southern England. *Quart. J. R. Met. Soc.*, **81**, p. 533.

O.M.I., 1955: *Definitions and specifications of water vapour in the atmosphere*, Publ. 79, Lausanne.

Öpik, E. J., 1951: Collision probabilities with the planets and the distribution of interplanetary matter. *Proc. Roy. Irish Acad.*, A, **54**, p. 165.

— 1955: Cosmic sources of deep-sea deposits. *Nature*, **176**, p. 926.

Packer, D. M., and Lock, C., 1951: The brightness and polarisation of the daylight sky at altitudes of 18,000 to 38,000 ft. *J. Opt. Soc. Amer.*, **41**, p. 473.

Paton, J., 1949: Luminous night clouds. *Met. Mag.*, **78**, p. 354.

— 1951: Simultaneous occurrence of aurora and noctilucent clouds. *Ibid.*, **80**, p. 145.

— 1954: Direct evidence of vertical motion at about 80 km provided by photographs of noctilucent clouds, p. 31, Proc. Toronto Met. Conf., *Roy. Met. Soc.*, London.

Penndorf, R., 1954: The vertical distribution of Mie particles in the troposphere. *Geophys. Res. Paper* 25, G.R.D., Cambridge, U.S.A.

Pettersson, H., and Rotschi, H., 1950: Nickel content of deep-sea deposits. *Nature*, **166**, p. 308.

Piotrowski, S. L., 1952: The collisions of asteroids. *A. J.*, **57**, p. 23.

Rakipova, L. P., 1947: Possible effect of dust on vertical air movements and on isothermy in the stratosphere. *Izv. Akad. Nauk SSSR*, **11**, p. 15 (translation T 199R, Defence Sci. Inf. Service, Canada, 1956).

Rense, W. A., Jackson, J. M., and Todd, B., 1953: Measurements of the inner zodiacal light during the total solar eclipse . . . *J. Geophys. Res.*, **58**, p. 369.

Schaefer, V. J., 1955: The question of meteoric dust in the atmosphere, unpublished.

Scorer, R. S., 1951: Billow clouds. *Quart. J. R. Met. Soc.*, **77**, p. 235.

Scultetus, H. R., 1949: Ältere Beobachtungen von Leuchtstreifen. *Z. f. Met.*, **3**, p. 272.

Spangenberg, W. W., 1949: Über die leuchtenden Nachtwolken, 1932—41. *Wetter und Klima*, **2**, p. 15.

Størmer, C., 1933: Height and velocity of luminous night clouds observed in Norway, 1932, Publ. 6, Univ. Obs., Oslo.

— 1935: Measurements of luminous night clouds in Norway, 1933 and 1934. *Astrophys. Norvegica*, **1**, No. 3.

— 1948: Mother-of-pearl clouds. Weather, **3**, p. 13.

Thomsen, J. W., 1953: The annual deposit of meteoritic dust. *Sky and Telescope*, **12**, p. 147.

Van de Hulst, H. C., 1947: Zodiacal light in the solar corona. *Ap. J.*, **105**, p. 471.

Vestine, E. H., 1934: Noctilucent clouds. *J. Roy. Astr. Soc. Canada*, p. 249.

Wexler, H., 1950: Annual and diurnal temperature variations in the upper atmosphere. *Tellus*, **2**, p. 262.

Whipple, F. J. W., 1930: The great Siberian meteorite and the waves, seismic and aerial, which it produced. *Quart. J. R. Met. Soc.*, **56**, p. 287.

— 1934: On phenomena related to the great Siberian meteorite. *Ibid.*, **60**, p. 505.

Whipple, F. L., 1943: Meteors and the earth's upper atmosphere. *Rev. Mod. Phys.*, **15**, p. 246.

— 1950: The theory of micro-meteorites, I, *Proc. Nat. Acad. Sci., Wash.*, **36**, p. 687.

— 1951: The theory of micro-meteorites, II. *Ibid.*, **37**, p. 19.

Whipple, F. L., 1952: Results of rocket and meteor research, *Bull. Amer. Met. Soc.*, **33**, p. 13.

— 1954: Photographic meteor orbits and their distribution in space, *A. J.*, **59**, p. 201.

— 1955: A comet model. III. The Zodiacal light. *Ap. J.*, **121**, p. 750.

— and Hawkins, G. S., 1956: On meteors and rainfall. *J. Met.*, **13**, p. 236.

Wyatt, S. P., and Whipple, F. L., 1950: The Poynting-Robertson effect on meteor orbits. *Ap. J.*, **111**, p. 134.

Zacharov, I., 1952: Influence des Perséides sur la transparence atmosphérique. *Bull. Astr. Inst. Czecho.*, **3**, p. 82.

Noctilucent Cloud Observations

By GEORG WITT, International Meteorological Institute in Stockholm

(Manuscript received May 15, 1957)

Abstract

Recent observations of noctilucent clouds from Swedish stations are presented, together with plans for the future observational programme. The results are discussed from the view-point of the rôle of meteoric dust in the formation of these clouds. Preliminary polarisation measurements, carried out in order to obtain an information about the sizes of the cloud particles, are described.

The question of the origin and composition of noctilucent clouds is still an unsolved problem in the physics of the upper atmosphere. These clouds are located at a level which is inaccessible for direct measurements and therefore, systematic visual observations and optical measurements are among the best methods for obtaining information about the phenomenon. Such observations have been carried out during the summers 1954, 1955 and 1956 at Torsta in Jämtland in NW Sweden, a place selected because of its favourable geographical position and extremely clear air. The results of the work during 1954—55 are described and discussed in a recent contribution by F. H. Ludlam who also presents an extensive review of the literature on this subject and gives a discussion of various theories forwarded in order to explain these cluds. The purpose of the present paper is to present the results of observations made at Swedish stations during the summer 1956, work considered as first phase of an observational programme on a broader scale for the year 1957. The work includes a set of measurements of polarization of the light scattered by the clouds aimed to yield additional information about the size of the cloud particles. The visual observations were made from two stations, Torsta near Östersund (63, 3N, 14, 6E) and Stockholm (59, 3N; 18, 1E). The work included continuous watching of the night sky during a period extending from the summer solstice to the end of August. In addition, occasional observations were made by fighter pilots of the Royal Swedish Air Force during night flights. The continuous recording of the presence as well as the absence of the clouds is in fact necessary especially with regard to the theory of a correlation of noctilucent clouds with meteor activity. Fig. 1 shows the distribution of the displays during 1956 together with a record of solar and ionospheric activity as reported from the astronomical observatory in Saltsjöbaden. The apparent coincidence between these latter events and noctilucent cloud displays may, however, be accidental and the figure is therefore presented without comment.

The position of the sun during the periods when noctilucent clouds were observed is shown in fig. 2.

The Observations

In the description we shall adopt the morphological classification introduced by Grishin (1955) and distinguish between four main cloud forms, i.e. nebulae or flares, bands, billows and eddies. For a more extensive description of these types see Ludlam (1957). The time will be noted in Swedish Mean Time, one hour in advance of GMT; the elevations are measured with a simple clinometer with an accuracy of $\pm 0{,}5°$.

July 2. Torsta: No observations possible because of low depression of the sun.

Stockholm: Faint display with bluish-white billows, first observed around 2330 between the azimuths 000—015 at 5—8 deg. elev. Except for changes in shape and intensity no motion was observed. The display vanished around 0115 drowned by the increasing background illumination. The date of this observation is remarkable with regard to the assumption of a maximum in the record of observed

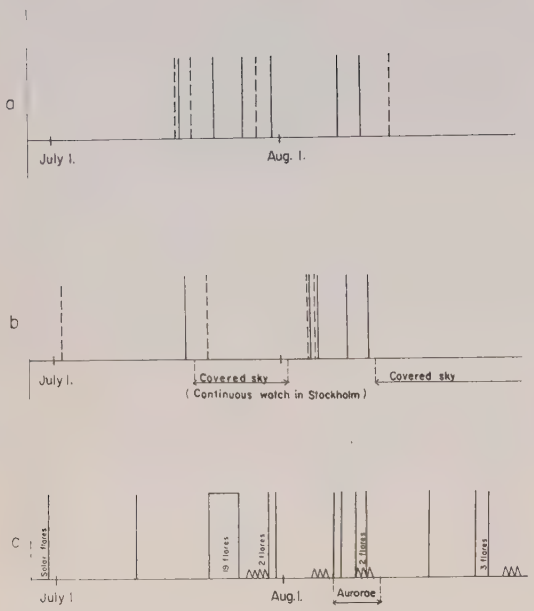


Fig. 1 a. Observations of noctilucent clouds from Torsta.
Dashed lines: 1954.
Solid lines: 1955.
b. Observations during 1956.
Dashed lines: Stockholm.
Solid lines: Torsta.
c. Reported solar activity during the summer 1956. Saw-tooth-shaped lines mark periods of ionospheric disturbances,

displays around the end of June due to increased meteoric activity (Vestine, 1934).

July 19. Torsta: First signs observed at 2340 in direction 005 through the last remainders of a dissipating Sc layer. The display consisted of pale blue, very sharp and almost straight bands parallel to the horizon at elevations 11—19 deg., with the most distinct parts at 11°, 12° and 16° resp. In addition there were some very faint billows in the north. Although the intensity increased somewhat after midnight, the brightness was very low and the clouds could only be observed by a trained eye. The extremely weak contrast made it impossible to photograph the clouds. The display vanished at 0043. Under the assumption that the average height of the clouds was 80 km the distance between two consecutive bands was estimated to be about 35 km. From Stockholm no observations were made.

July 22. Torsta: No observations because of cloudy weather.

Stockholm: Clouds first observed at 2325, consisting of billows at 12° elev. in direction N together with diffuse, fibrous masses down to the horizon. At 0105 the clouds covered a range of 45 deg. of azimuth with the center in NNE. In addition a thin, dust-like nebula appeared above the billows around 0100 which remained visible until 0120 when a rapidly spreading Sc layer made further observations impossible. The clouds had the typical bluish-white colour, the brightness culminating around 0030, and appeared to move from East to West.

August 5. Torsta: The characteristic fibrous bands of bluish colour were first observed at 2330 in the clear space between an Sc layer and the horizon. Occasional breaks in the Sc cover revealed the existence of noctilucent clouds over the entire horizon between 330° and 000° azimuth. The elevations of the brightest bands were 5°, 6° and 7° respectively. After midnight, when the Sc slowly disappeared in the north a very intense display was revealed with the brilliance culminating around 0115. At this time a delicate bluish white shining nebula or flare was visible over the northern horizon with elevations up to 30°, the center being located in the vicinity of Capella. In addition a system of long stretched fibrous bands reminding of "polar bands" was seen in NNE, the orientation of the bands indicating a motion from N—NE. The bands were "connected" by smaller billows which gave the entire system the shape of a cobweb. After its culmination the system slowly faded away without significant changes of shape. At 0200, however, a new series of billows appeared at very high elevations in the east. The contrast of these clouds was rather low and they vanished already at 0210. These billows were also observed from Stockholm. (see below).

Stockholm: The first clouds were observed at 2350, extending from 330° to 030° azimuth with the maximum elevation of 6°. The details of the display were not very marked. In the course of the time, however, the intensity increased and about 0200 typical billows became visible in NNE at 13° elev. Below 5° the morning glow covered the display. At 0215 a thin nebula was seen at unusually high elevations, reaching up to the

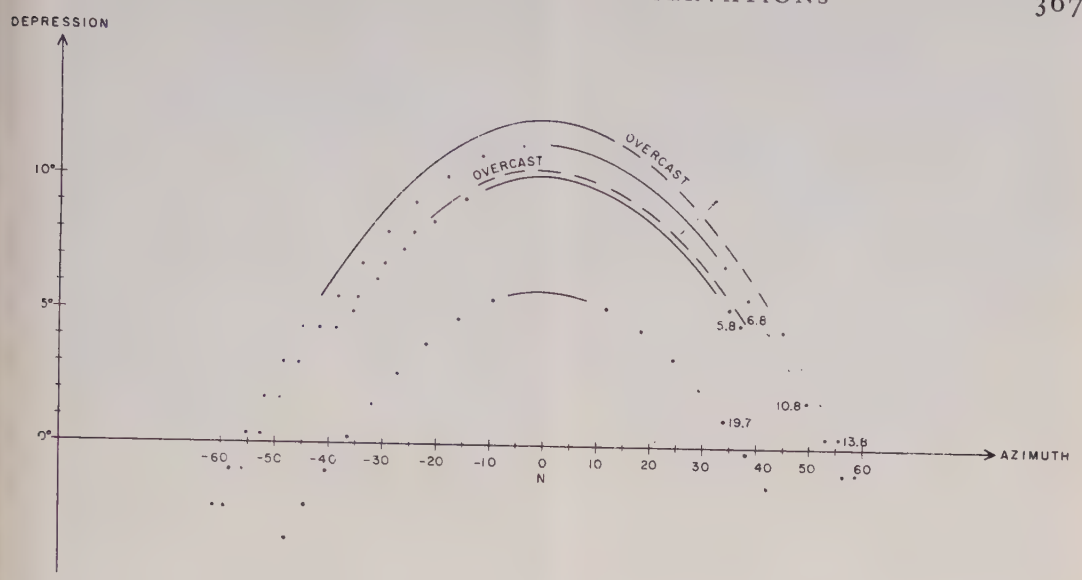


Fig. 2. Position of sun during NIC displays 1956.
The periods during which the clouds were visible are marked by the solid curves.
The dashed parts denote overcast weather.

polar star (59°). This nebula gave the impression of a faint aurora with a slight bluish tinge. The last observation was made at 0230.

August 6. Torsta: Silvery-blue bright spots observed over the horizon in WNW at 2240. Shortly afterwards the characteristic billows appeared in NW and N with elevations up to 15°. A slowly developing Ac system made it impossible to make observations after 2310. There were no traces of noctilucent clouds towards NE.

Stockholm: The northeastern sky was covered by Ac and Ci. A weak display appeared about 2145 over the northern horizon with a very low brilliance and diffuse in structure. The clouds could first at 2220 be identified as NLC. The maximum elevation was about 8° and the display seemed to extend from 330° to 030° azimuth. After 2330 the display became even weaker and disappeared completely about 0030.

Comparing the maximal elevations measured this night at Torsta and Stockholm respectively we can assume that the observed clouds did not extend farther south than about 66° N.

August 10. Torsta: Although this evening was one of the few during the past summer when the sky was completely free from clouds, no noctilucent clouds were observed before midnight. Immediately over the horizon the sky was covered by a haze layer with an intense red illumination, showing several haze bands. At the upper edge of this layer in NNE thin bluish bands appeared some minutes after midnight. The bands which had elevations up to 6° only, could later be identified as noctilucent clouds. The brightness was very low owing to the long distance to the clouds, estimated under the assumption of an average height of 80 km to be about 800—900 km toward NNE. In addition to the fibrous bands faint but distinct billows could be seen, giving the vague impression of a slow movement from NNE. The clouds vanished at 0215. It may be interesting to notice that during the period between the 9/8 and 14/8, this night was the only one when no aurora could be observed.

Stockholm: The clouds were not seen from Stockholm. This fact together with the very low elevations measured at Torsta indicate again that the cloud system probably did not extend to latitudes below 67°—68°.

August 13. Torsta: This night the most outstanding display of this year could be observed, compensating for the more than unfavourable weather conditions which spoiled a major part of the summer. The first bluish billows could

be seen in NW at 25° elev. already at 2150. These billows formed the upper edge of a layer of diffuse noctilucent clouds extending from 290° NW towards NNE where the sky was covered by Sc. The development of the display was photographed with a Paillard Bolex movie camera with an approximate rate of three frames per minute. A sequence of 8 pictures with 5 min. intervals, made from the time-lapse film is shown in fig. 4 The colour of the clouds was silvery-blue and the brilliance of the display unusually high. After midnight the brilliance decreased and except for a few weak stripes, the only sign of a display, seen in NNE, was the unusually bright twilight. At 0115 a faint aurora was seen for a short while in the north, for the first time in connection with noctilucent clouds. This night unfortunately no observations were made from Stockholm. Air Force pilots reported, however, that noctilucent clouds were observed during two flights at latitude 54°—60° N between 1900 and 2200 SMT. During a flight at 9 000 m the clouds were seen in the sector 330°—015° at elevation 15°. From the second flight at 7500 m the clouds were observed in the sector 320°—340° at approximately 6—7 degrees elevation. The height of the cirrus level was as reported 7,000—10,000 m. This ends the observations for 1956. The remaining period was characterized by a continuous cloudy weather which made further work impossible.

The observations described above show two features that we should like to emphasize. Both are of interest from the point of view of a theory forwarded by Vestine a.o. (see Ludlam) suggesting that the origin of the clouds should be sought in an inflow of particles from outside the earth's atmosphere. To begin with we find that our observations fall within a period of increased meteoric activity (cf. fig. 3) which at least does not contradict the theory mentioned. Since we have to consider the inflow of small particles we do not necessarily need to find a correlation between noctilucent cloud displays and the dates of maxima intensity of the actual meteor showers. Any discussion of the statistical significance will, however, require further series of observations. It would be advisable that such observations should be carried out at stations with similar illumination conditions and include a recording of the presence and the absence of the clouds as well as unfavourable viewing conditions. We hope that such observations will be made at various stations during the coming Geophysical Year.

There are furthermore indications that the clouds are really confined to higher latitudes. This fact may of course be an effect of special meteorological conditions during the summer-season at the particular level where the clouds are located. Assume, however, that the cloud particles originate from the interplanetary space, swept up by the earth. Assume further that such particles carry electric charges. Then according to the theory of motion of a charged particle in the magnetic field of the earth, the inflow of such particles will be latitude-dependent provided the magnetic rigidity of the particles does not exceed a certain limit. (Störmer 1955). Let us as an example consider a particle with radius $a = 10^{-6}$ cm and density 2 g/cm^3 having a velocity of 30 km/s relative to the earth and carrying Z elementary charges. Let us for simplicity neglect the influence of the gravitational field which is justifiable for particles of this size. We then have the magnetic rigidity

$$H\varrho = \frac{mvc}{Ze}$$

$m = 2 \times 4\pi \cdot a^3/3$ denotes the mass of the particle, v the relative velocity, $c = 3.10^{10}$ cm/s and $e = 4.8 \cdot 10^{-10}$ esu. Inserting the assumed numerical values in the formula above we obtain

$$H\varrho = 15.7 \times 10^8/Z \text{ gausscm}$$

Now, following Störmer's discussion of the orbits of charged particles we find that the cut-off magnetic rigidity for arrival near the geomagnetic equator is approximately 3.3×10^7 gausscm. Comparing this value with our result above we see that a particle carrying more than 48 elementary charges will never arrive at the equator. If our particle is assumed to be spherical and thus the electrostatic capacity = the radius a, 48 elementary charges correspond to a potential of 6.9 volts, which should be a reasonable value for a particle in the interplanetary space. We see thus that at least for the smallest particles a latitude dependence in the inflow can not be excluded. This again points toward the necessity of world-wide observations, if possible also from the southern hemisphere.

Fig. 4. Noctilucent clouds over the North-Western horizon, photographed from Torsta on August 13, 1956. The sequence of frames, made from a 16 mm time-lapse film shows the development of the display in 5 min. intervals from 2210 to 2250 Local Mean Time. (Film: Kodachrome, exposure 5 secs with opening 1,4, f = 250 mm) The right-hand edge of the frames points toward North.
A direct comparision of the features in the pictures is difficult since, unfortunately, the control over the time of exposure and the direction of the camera was not satisfactory (See the difference between the last two frames).

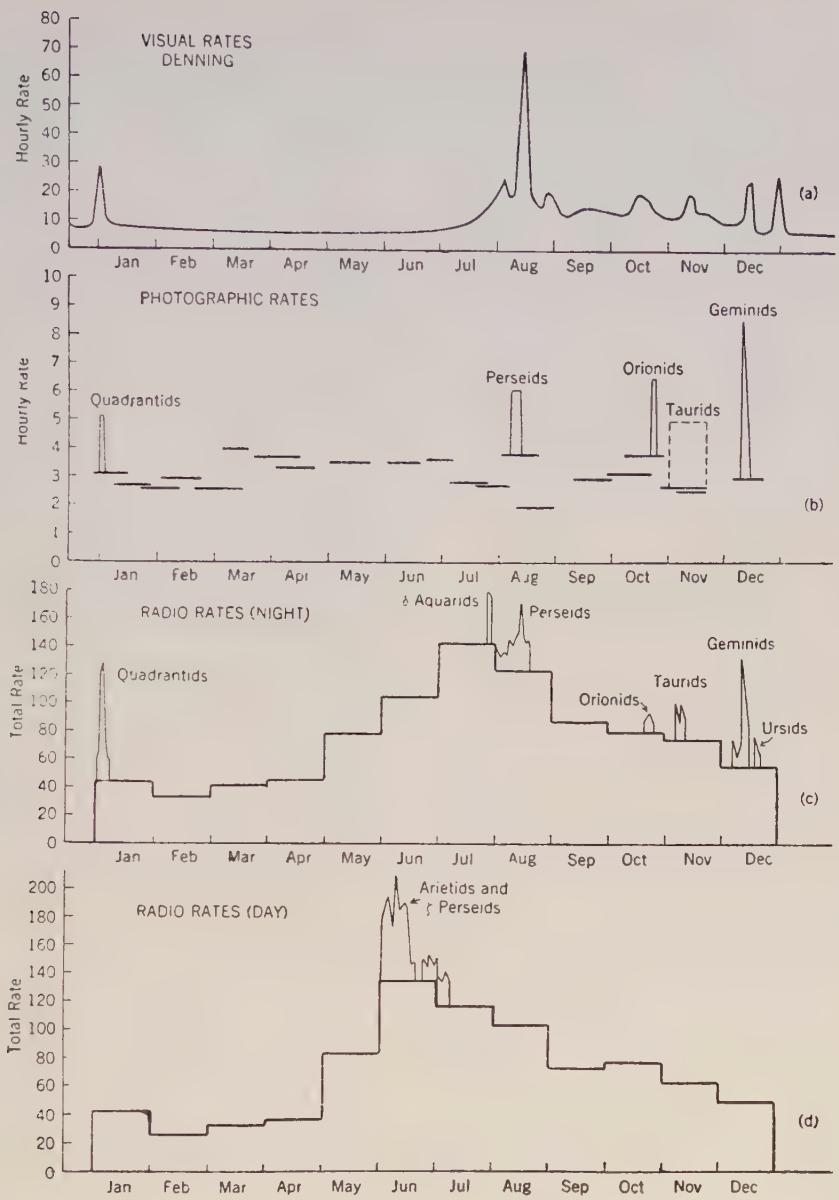


Fig. 3. Variation of inflow of visible, photographic and radar meteors with season. (Whipple & Hawkins, 1956).

Polarization Measurements

The abscence of emission lines from the spectrum of noctilucent clouds, and the increase of intensity of the spectrum towards shorter wavelengths indicate that we have to deal with reflected sunlight. The fact that this light has been found to be partly polarized shows furthermore that the sunlight is scattered by particles of the order of magnitude comparable to the wavelength of the visible light, or smaller. From Mie's theory of scattering of light in turbid media (Mie 1908)

it follows that in this size range the variation of polarization with the scattering angle and wavelength is characteristic for particles with given radii. Based on this theory a set of preliminary polarization measurements has been made at Torsta during the summer 1956 with the aim to obtain additional information about the size of the cloud particles. For our purposes we consider the clouds as an infinitely dilute suspension of spherical scatterers. The particles are illuminated by the (parallel) rays of the sun consisting of two incoherent components vibrating parallel and perpendicular, respectively, to a plane defined by the sun, the scatterer and the detector. The intensities of the two components of the scattered light are, according to Mie, expressed by the formulae

$$I_1 = \frac{\lambda^2}{4\pi R^2}\left|\sum_{\nu=1}^{\infty}\left\{\frac{a_\nu}{\nu(\nu+1)}\Pi_\nu(x) + \right.\right.$$

$$\left.\left. + \frac{p_\nu}{\nu(\nu+1)}\left[x\cdot\Pi_\nu(x) - (1-x^2)\,\Pi_\nu'(x)\right]\right\}\right|^2$$

$$I_2 = \frac{\lambda^2}{4\pi R^2}\left|\sum_{\nu=1}^{\infty}\left\{\frac{a_\nu}{\nu(\nu+1)}\left[x\cdot\Pi_\nu(x) - \right.\right.\right.$$

$$\left.\left.\left. - (1-x^2)\,\Pi_\nu'(x)\right] + \frac{p_\nu}{\nu(\nu+1)}\Pi_\nu(x)\right\}\right|^2$$

Here λ denotes the wavelength of the light, R the distance between scatterer and detector and $x = \cos\vartheta$ where ϑ denotes the scattering angle. The coefficients a_ν and p_ν are functions of the refractive index n and the parameter $\alpha = \frac{2\pi r}{\lambda}$ where r is the radius of the particle. $\Pi_\nu(x) = \frac{\partial P_\nu(x)}{\partial x}$ and $\Pi_\nu'(x) = \frac{\partial^2 P_\nu(x)}{\partial x^2}$, where $P_\nu(x)$ denotes the Legendre polynomial of degree ν.

The degree of polarization is defined as the ratio between the two components of the scattered light. For small values of α the contribution from higher order terms in the series is negligible and the polarization will follow Raleigh's law. $\left(\frac{I_1}{I_2} = \frac{1}{\cos^2\vartheta}\right)$. With increasing radius or decreasing wavelength the deviation from this formula becomes more and more pronounced. (Numerical values can be obtained e.g. from Appl. Math. Ser., 4). Thus we can expect that a study of the variation of the polarization with the scattering angle and the wavelength will give an information about the presence of relatively large particles.

The instrument used for the first experiments at Torsta consisted of a type 931 photomultiplier tube, and interference filter with $\lambda = 4950$ Å, and a turnable Polaroid filter. As a condensor we used a Fresnel lens made of Perspex with $f = 30$ cm giving the instrument an opening angle of 5° in the vertical and 3.5° in the horizontal. The intensity was read from a light-spot galvanometer at every fifth degree of azimuth with the instrument kept at constant elevation. The relation between the scattering angle ϑ, the depression of the sun Θ, the elevation of the instrument γ and the azimuth relative to the sun λ used for the evaluation of the results is obtained by solution of the spherical triangle ABD in fig. 5.

$\cos(\pi - \vartheta) = \cos\lambda\cdot\cos\Theta\cdot\cos\gamma - \sin\Theta\cdot\sin\gamma$

The difficulty in the photometry of the twilight sky lies in the fact that the intensity of the light is very low and the measurements are disturbed by the presence of light not due to direct scattering. This is especially true when, as in our method, the light is collected from a relatively large area of the sky, out of which only a fraction is covered by noctilucent clouds. As a consequence of this a measurement requires a bright and extensive display. Also, since at small scattering angles the polarization is low, the accuracy of the measurements will be poor. We are at the present making experiments in order to refine our method by using a stereo camera with a pair of polaroid filters with the axes of orientation mutually perpendicular. The filters are

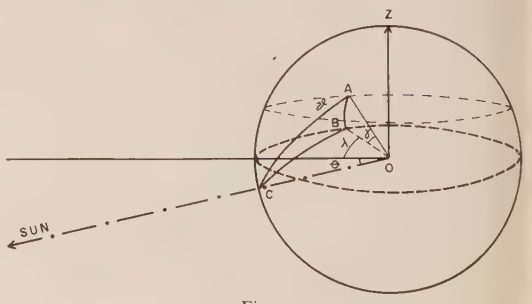


Fig. 5.

mechanically coupled with the photometer by the aid of which we can find the approximate inclination of the plane of polarization. Chosing a suitable time of exposure the relative intensities can be evaluated directly on the clouds by photodensitometry.

During the last summer only one single occasion occurred when the viewing conditions permitted a measurement. This was the display on August, 13 when in addition the disturbing reddish yellow light from immediately above the horizon was screened by a layer of Sc. As a comparison the polarization of the clear sky was measured the 12 August in the same part of the sky and during the same time interval. The comparison indicated that without the use of a filter the values obtained from the clear sky only slightly exceeded the values from noctilucent clouds. Both measurements showed, however, a marked departure from Raleigh's law thus indicating the presence of not directly scattered light. Insertion of the green filter gave as a result an increase of the values for the clear sky while in the case of noctilucent clouds we found a pronounced decrease. (See fig. 7.) Even if we are not yet permitted to draw definite conclusions we can consider this as an indication of the presence of particles with radii of the order of magnitude of a few tenths of microns giving a predominating contribution to the scattered light in accordance with the results obtained from spectroscopic investigations. Nothing can, however, be stated about the relative number of these large particles since in the size range in question the intensity of the scattered light varies with the fifth power of the radius.

Even though our measurements are too scarce to permit a discussion we wish to present the results with the hope that measurements of this kind will be made also at other stations during the Geophysical Year. This would be especially valuable with regard to the fact that displays with sufficient brightness are relatively rare and coincide only occasionally with cloudless sky.

Acknowledgements

The work was carried out as a part of the Cloud Physics Research Program of the Institute of Meteorology, University of Stockholm and was supported by Statens Tekniska Forskningsråd. We wish to express our thanks to staff members of the Institute and in particular Mr. Dusan Djuric for his valuable assistance during the summer; to Dr. N. Herlofson and Mr. E. Å. Brunberg of the Institute of Electronics at the Royal Institute of Technology for advices and discussion; and Dr. R. M. Goody, London, who initiated and supported our work.

REFERENCES

Grishin, N. I., 1955: On the Structure of Noctilucent Clouds, *Met. i Gidr.*, No. 1, p. 23 (in Russian).

Ludlam, F. H., 1957: Noctilucent Clouds, *Tellus*, 9, No 2, p. 000.

Mie, G., 1908: A contribution to the Optics of Turbid Media . . ., *Ann. Physik*, Ser. IV, **25**, pp. 377—445 (in German).

Vestine, E. H., 1934: Noctilucent Clouds. *J. Roy. Astro. Soc.*

Whipple, F. L., Hawkins, G. S., 1956: *Journal of Met.*, **13**, No 3.

Störmer, C., 1955: *The Polar Aurora*, Oxford University Press.

Tables of Scattering Functions for Spherical Particles, Nat. Bureau of Standards, *Appl. Math. Ser.*, **4**, Washington, 1949.

On a Special Aspect of the Condensation Process and its Importance in the Treatment of Cloud Particle Growth

By C. ROOTH, International Meteorological Institute in Stockholm

(Manuscript received May 10, 1957)

Abstract

Available data on the rate of exchange of molecules between liquid water and vapour in contact with it are put in relation to atmospheric condensation and evaporation processes. It is shown that this exchange process excerts a quenching action on the rate of growth of small droplets that should contribute to the formation of a broad size spectrum. Surface films may enhance the effect considerably. Experimental data are needed to decide finally the meteorological significance of the effect.

In theoretical treatments of cloud particle growth it is generally assumed that if the chemical composition as well as the size and the surface temperature of a droplet were known, the equilibrium value of water vapour pressure determined by these data can be used as a boundary condition for the diffusion of water vapour to or away from the droplet studied. This is a reasonable assumption as long as the net transport of water vapour is small as compared to the rate of exchange of molecules across the surface of the droplet. It is the purpose of this presentation to show, that this is not always the case, and to indicate some consequences of this fact.

The exchange process at the liquid-vapour boundary may be split into two processes, namely the escape of molecules from the liquid into the vapour phase and the capture of vapour molecules at the liquid surface.

If q_e denotes the average rate at which water molecules leave the liquid phase, given in g/sec/cm^2, and if q_c represents the rate at which molecules from the vapour hit the liquid surface, then at equilibrium:

$$q_e = \alpha q_c \qquad (1)$$

where α is the condensation coefficient that indicates how large a proportion of the molecules hitting the surface are actually caught. According to measurements by ALTY & MACKAY (1935) α for water has a value of 0.036 at 10° C, decreasing somewhat with rising temperature. Alty also states that for vapours in contact with a crystalline phase of the same substance, the condensation coefficient is near to 1 if the molecules do not have a dipole moment, but it is considerably lower for polar molecules. No value for ice is reported by him.

The magnitude of q_e can be estimated in the following way:

According to gas kinetic theory (cf KENNARD (1939))

$$q_c = \frac{e}{(2\pi R T)^{1/2}} \qquad (2)$$

is the rate at which vapour molecules hit the surface. Here e is the vapour pressure in dynes/cm^2 immediately at the surface (i.e. at a distance from it approximately equal to the mean free path of the water molecules in the

air. R is the gas constant per gram of water vapour, and T is the absolute temperature.

At equilibrium equation (1) will be valid so that

$$q_e = \frac{\alpha e_0}{(2\pi R T)^{1/2}} \tag{3}$$

Now q_e as well as α are independent of the state of the vapour phase so that any departure from equilibrium conditions in the exchange of molecules across the surface must be accompanied by a departure of q_c from the value indicated by (1). This departure is in fact seen to be proportional to the ratio of the actual rate of evaporation or condensation to the exchange rate q_e.

It is thus clear that when treating the problem of diffusion of water vapour to or from a liquid water surface for which the equilibrium pressure, e_0, of water vapour is known, one may not apply the boundary condition that.

$$e = e_0 \tag{4}$$

without qualification, but that a term proportional to the rate of evaporation should be included, giving:

$$e = e_0 \left(1 - \frac{q}{q_e}\right) \tag{5}$$

where q denotes the rate of evaporation and q_e is defined above.

Making use of Alty's value for the condensation coefficient one finds that q_e at 10° C is equivalent to a rate of evaporation of 180 mm/h, a value so far above what is normally found in nature that the condition (4) can be used safely when treating the evaporation from lakes and other flat surfaces. The significance of the difference between (4) and (5) may, however, be judged more easily by the following argument: If one applies the boundary condition (2) and studies conditions at a small distance s from the liquid surface under the assumption that the transport of water vapour through the intervening layer takes place only through molecular diffusion, then one arrives at the following expression for the vapour pressure e_s:

$$e_s = e_0 - \frac{qs}{K} \tag{6}$$

where K is the appropriate coefficient of diffusion. Comparing (5) and (6) it is clear that the use of the boundary condition (4) in a problem of evaporation or condensation is equivalent to the neglect of an air layer through which the vapour has to pass by molecular diffusion. The thickness s of that layer is given by the relation

$$s = \frac{Ke_0}{q_e} = \frac{K}{\alpha}(2\pi R T)^{1/2} \tag{7}$$

Again, using the values appropriate at 10° C and 1,000 mb, we find that $s = 5.1\,\mu$ which can be safely neglected in many meteorological applications. However, in dealing with cloud droplets this is not so.

The growth equation normally used for cloud droplets may be written:

$$\frac{dm}{dt} = 4\pi K r (e_a - e_0) \tag{8}$$

where m is the mass of the droplet, K is the same coefficient of diffusion as was used in (6), r is the radius of the droplet, e_a is the ambient partial pressure of water vapour, and e_0 is the equilibrium value determined by the temperature, the radius of curvature, and the chemical composition of the droplet (cf. HOWELL (1949) or SQUIRES (1952)). The actual form of e is immaterial for this presentation.

According to the reasoning above, e_0 in (8) should be replaced by

$$e_0 \left(1 - \frac{q}{q_e}\right)$$

with

$$q = -\frac{1}{4\pi r^2}\frac{dm}{dt} \tag{9}$$

the net rate of transport of water vapour per unit area of the droplet surface.

Making use of (7) we finally get:

$$\frac{dm}{dt} = 4\pi K r (e_a - e_0)\frac{r}{s + r} \tag{10}$$

A comparison between the two expressions for the growth rate is made in fig. 1 where the straight line represents equation (8) and the curved line below it represents equation (10). The value of s is taken to be 5 μ, applying

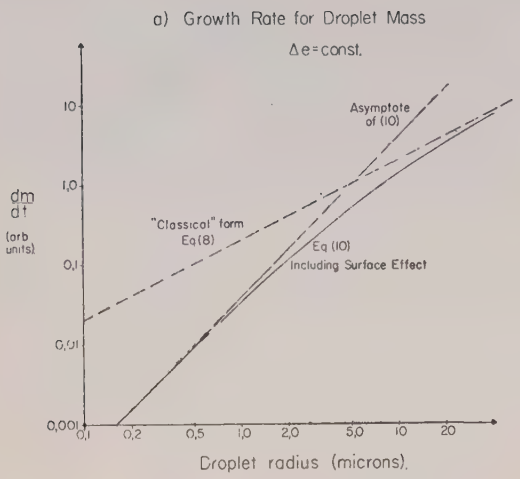
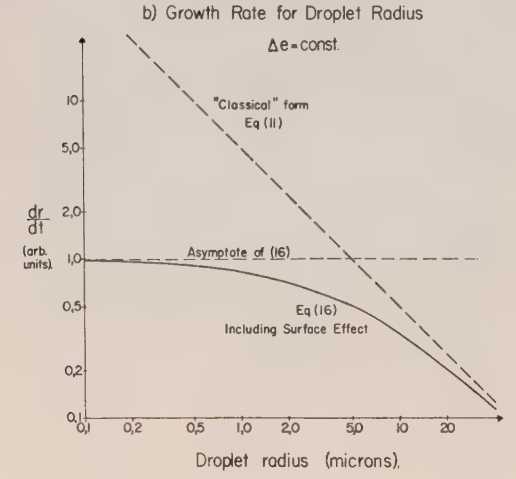


Fig. 1. a) Comparison of theoretical growth rates for a given value of $e_a - e_0$ according to equations (8) and (10).



b) The same growth rates transformed into rates of change of droplet radius. (Equations (11) and (16)).

at 10° C and 1,000 mb. No conclusions as to the range of s can be drawn at the present time, since the information regarding the variation of α is incomplete. If, however, α were independent of the temperature, then s according to (7) would increase somewhat with a rise in the temperature.

However spectacular the difference displayed in fig. 1 may be, it is still difficult to judge its significance since the actual values of the excess humidity may vary considerably between the different members of a cloud droplet spectrum.

In order to elucidate this point one might look upon the development of the particle spectrum in a rising air column below the cloud base level. Hygroscopic particles kept in an environment of constant relative humidity will take on their equilibrium size according to the theory by KÖHLER (1936). Equilibrium curves relating droplet radius to relative humidity for various nucleus sizes are shown in fig. 2. KEITH and ARONS (1954), among others, have shown that the smaller particles follow their equilibrium curves almost perfectly, while the larger ones may lag behind considerably in their growth. Their data were gained by an integration of the usual equation of growth for the individual particles, following them in their motion, and were checked by experiments. The droplets that they used for the measurements of growth rates were however too large to show the effect discussed here.

A simple approximation to the integrated time lag of the particle growth may be reached by the following argument:

Equation (8) can be written as

$$4\pi\varrho r^2 \frac{dr}{dt} = 4\pi k r\,(e_a - e_0) \qquad (8a)$$

where ϱ is the density of liquid water. Thus

$$\frac{dr}{dt} = \frac{K}{\varrho r}(e_a - e_0) \qquad (11)$$

If now r_a is the equilibrium radius corresponding to the ambient vapour pressure e_a, then according to the mean value theorem:

$$e_a - e_0 = -(r - r_a)\frac{\partial e_0}{\partial r} \qquad (12)$$

where the partial derivative should be determined for a droplet radius lying inbetween r and r_a. If the departure from equilibrium is not too great a sufficient approximation is reached by choosing the value r. Equation (11) may now be rewritten in the form:

$$\frac{d(r - r_a)}{dt} = -\frac{K}{\varrho r}\frac{\partial e_0}{\partial r}(r - r_a) - \frac{dr_a}{dt} \qquad (13)$$

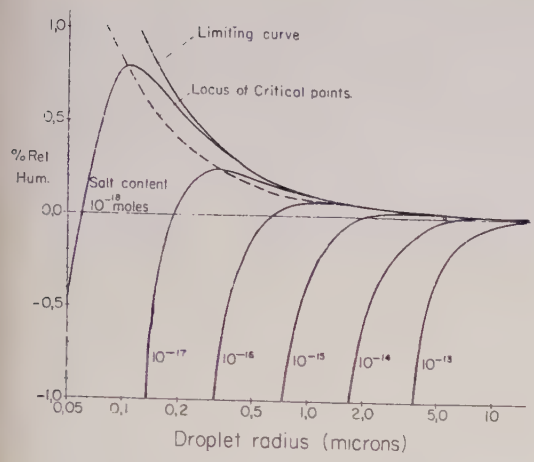


Fig. 2. Equilibrium curves for hygroscopic nuclei. The equilibrium is stable to the left and unstable to the right of the critical point.

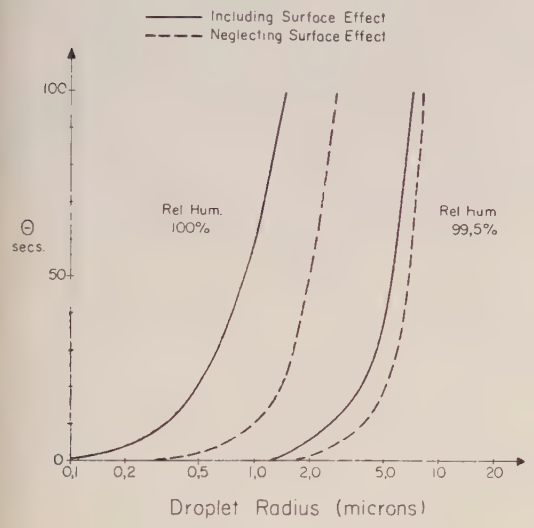


Fig. 3. Comparison of theoretical lag constants for cloud droplets of equilibrium size at 99.5 and 100 % relative humidity. The data apply at a temperature of 10° C.

Denoting the coefficient of $(r-r_a)$ in the right hand member of (11) by Θ^{-1}, so that

$$\Theta = \frac{\varrho r}{K\dfrac{\partial e_0}{\partial r}} \tag{14}$$

two simple interpretations of equation (13) may be given. If the ambient vapour pressure is constant, then small deviations from the equilibrium radius will decay exponentially with a time constant equal to Θ. If, on the other hand, e_a undergoes a slow and steady variation, then the two terms in the right hand member of eg (13) will balance approximately. One finds in this case that the actual value of r will lag behind r_a so that at each instance it will be approximately equal to the r_a-value valid Θ seconds earlier.

It is easily seen, that the same argument could be developed starting with equation (10), i.e. the modified form of the growth equation, producing the result that

$$\Theta = \frac{\varrho(r+s)}{K\dfrac{\partial e_0}{\partial r}} \tag{15}$$

These values for the lag Θ of the droplet growth should be quite accurate as long as the value of Θ does not vary appreciably over time intervals comparable with Θ itself.

In fig. 3 two sets of curves are shown, relating droplet radius to lag constant for droplets in equilibrium with an environment of 99.5 % and 100 % relative humidity respectively. The stippled lines represent the values gained when using the equation in the normal form, while the fully drawn curves indicate the deviations caused by the correction factor $(1+s/r)$. When looking at these curves one should bear in mind that values of Θ around 100 seconds or more will not represent actual lags except possibly in cases of extremely slow cooling, and that lag values around 1 second or less have little significance as long as one is not working with very rapid expansions, such as in an Aitken counter. These considerations do, however, still leave room for a considerable influence of the correction factor presented here in the range of particles that have critical radii around a few microns.

An argument similar to that of HOUGHTON (1938) provides still another presentation of the difference between the two versions of the growth equation discussed here. Starting with equation (11)

$$\frac{dr}{dt} = \frac{K}{\varrho r}(e_a - e_0) \tag{11}$$

and the corresponding modified form, which it seen to be

$$\frac{dr}{dt} = \frac{K}{\varrho(r+s)} (e_a - e_0) \qquad (16)$$

and treating s as a constant we may integrate the equations to give the following alternative expressions:

$$r_2^2 - r_1^2 = \frac{2K}{\varrho} \overline{(e_a - e_0)} (t_2 - t_1) \qquad (17)$$

and

$$(r_2+s)^2 - (r_1+s)^2 = \frac{2K}{\varrho} \overline{(e_a - e_0)} (t_2 - t_1) \qquad (18)$$

for the growth of a cloud particle. Here the bar denotes an average over the time interval considered. The only difference between Houghtons reasoning and the present argument is that $(e_a - e_0)$ here represents the deviation of the ambient vapour pressure from the equilibrium at the surface of the droplet, including the effect of temperature deviations between droplet and environment, while Houghton worked under the simplifying assumption that e_0 could be represented by the saturation vapour pressure at the ambient temperature of the environment. If to begin with, $\overline{e_a - e_0}$ is prescribed, in analogy again with Houghton, then we see that the ratio R between the time taken for a droplet to grow from a size r_1 to a size r_2 according to the corrected and the uncorrected formula is

$$R = \frac{r_1 + r_2 + 2s}{r_1 + r_2} \qquad (19)$$

For the growth from 5 to 10 μ this ratio is 5: 3, again accepting $s = 5\,\mu$. This change is quite notable as the time scale involved here will generally be tens of seconds up towards a minute. However, in the dynamic development of a cloud the time scale is prescribed rather than the relative humidity, as the latter will increase if the water vapour is not consumed at a rate corresponding to the rate of cooling.

The expected value of the maximum supersaturation attained in the initial cloud formation would under such circumstances be higher according to the modified theory than otherwise. That value is the limiting factor that determines the partition of the particle spectrum into one part that is growing actively and one inactive part. A higher maximum value of the supersaturation means a smaller limiting size to the nuclei growing actively. This leads to the somewhat surprising conclusion that an effect that slows down the growth of small particles under a given excess humidity may in fact by causing and increase in the relative humidity make it possible for more of the small particles to grow actively in the cloud.

It has thus been shown that the distribution of water vapour in the air in the vicinity of a drop that is evaporating or growing by condensation cannot be determined by considering the diffusion process alone unless the dimensions of the droplet are large compared to a distance s which is primarily determined by the rate of exchange of molecules between a water surface and the vapour phase under equilibrium conditions. If this distance s were of the order of 10 μ, then the results of any investigation regarding the development of cloud particle spectra would be seriously in error, unless the effect of the phase transition on the vapour pressure in the immediate vicinity of the droplets were included. If, on the other hand, s were around 1 μ or less, then the practical consequences in cloud physical problems would be very small, as droplets in the submicron range tend to adapt themselves so rapidly to changes in environmental conditions that a change of the time scale by a factor of 10 or even 100 is of little importance in that size range.

Whatever the scaleof s may be, its importance will be greater the more rapid the cooling is in the cloud, as its effect is one of increasing the lag time of the particle growth, leaving the expressions for the equilibrium conditions unchanged.

Now the actual value of $s = 5\,\mu$ used in the present discussion is subject to considerable uncertainty. It is based on the value of $\alpha = 0.036$ gained by ALTY & MACKAY (1935) by measurements of the rate of evaporation of drops suspended in water vapour of less than equilibrium pressure. The main difficulty in his experiment lies in the determination of the actual surface temperature. According to a discussion by VOLMER (1939) any systematic errors in the determination of α-values tend to produce results that are too low. The presence of any surface contamination is

particularly serious, but also the problems of heat and vapour diffusion are great when dealing with a substance of relatively high volatility such as water. If the true value of α were larger, then s would be proportionally smaller. Its effect would then soon approach the limit of meteorological insignificance. On the other hand, it is known that rain water contains organic matter. If that matter is present already in the cloud droplets, then it may cause the actual α-values to be much lower than for pure water. Consequently a wider spectrum of droplet sizes could be produced than in a cloud of uncontaminated water droplets. This effect would then greatly favour the onset of droplet growth by coalescence, especially in cases of relatively strong updraughts. All these conclusions will, however, remain speculations until more and better information regarding the exchange of water molecules across a vapour-liquid interface is made available, as well as data on the possible presence of surface films or cloud droplets.

Acknowledgements

Thanks are due to members of the International Meteorological Institute for valuable criticism and discussions. The study was made possible through the support given by Statens Tekniska Forskningsråd for work on cloud physics and evaporation problems.

REFERENCES

Alty, T. Mackay, C. A., 1935: The accommodation coefficient and the evaporation coefficient of water. *Proc. Roy. Soc.*, London, **A 199**, 104—116.

Houghton, H. G., 1938: Problems connected with the Condensation and Precipitation Processes in the Atmosphere. *Bull. Amer. Met. Soc.* **19**, 152—159.

Howell, W. E., 1949: The growth of cloud drops in uniformly cooled air. *J. Met.* **6**, 134—149.

Keith, C. H., Arons, A. B., 1954: The growth of sea salt particles by condensation of atmospheric water vapour. *J. Met.* **11**, 173—184.

Kennard, E. H., 1939: Kinetic theory of Gases. Mac Graw Hill, New York.

Köhler, H., 1936: The nucleus in and the growth of hygroscopic droplets. *Trans Faraday Soc.* **32**, 1152—1161.

Squires, P., 1952: The growth of cloud drops by condensation. *Australian J. of Sci. Res.*, **A 5**, 59—86.

Volmer, M., 1939: Kinetik der Phasenbildung. Verl. Th. Steinkopff, Leipzig.

On the Diffusion of Heated Jets

By F. H. SCHMIDT, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

(Manuscript received, October 13, 1956)

Abstract

A general expression is obtained for the propagation of a heated jet in a stagnant environment. It is shown that W. Schmidt's and Tollmien's solutions are contained in the general one. A new cross-sectional velocity distribution is proposed which seems to be preferable above the generally accepted Gaussian distribution.

1. Introduction

TOLLMIEN (1926) and W. SCHMIDT (1941) were the first to consider the turbulent subsonic flow of a fluid jet through a stagnant environment. Tollmien restricted himself to the case of an incompressible fluid of the same density as the environment, i.e. no buoyancy forces were taken into account. W. Schmidt extended Tollmien's theory to the case of a heated jet rising due to buoyancy. The relative density differences were connected with temperature differences.

Recently PRIESTLEY and BALL (1955) arrived at a solution that can be considered as a more general one, containing as it does both Tollmien's and Schmidt's results.

The authors do not seem to have recognized their solution as such a generalization, however. Moreover, it was obtained by the introduction of special assumptions as to the contribution of velocity and temperature in a plane perpendicular to the jet's axis. These assumptions include the same error function for both velocity and temperature profiles, which seems to be inconsistent with all experimental evidence.

It will be shown here that it is possible to arrive at a general solution without introducing any special assumptions about the velocity and temperature distributions in the jet. Afterwards a velocity profile will be proposed that does not show some of the disadvantages of profiles introduced so far.

2. Two-dimensional jet

First consider a two-dimensional jet, emerging from an infinitely long aperture and assume that density differences only play a role in as far as they effect buoyancy forces. The z-axis is taken vertical, the y-axis perpendicular to the aperture and molecular friction and conduction are ignored. Then the following equations hold for stationary two-dimensional flow:

$$\frac{\partial}{\partial z}(\overline{w}+w')+\frac{\partial}{\partial y}(\overline{v}+v')=0 \qquad (1)$$

$$\frac{\partial}{\partial z}(\overline{w}+w')^2+\frac{\partial}{\partial y}(\overline{w}+w')(\overline{v}+v')= -\frac{1}{\varrho}\frac{\partial p}{\partial z}-g \qquad (2)$$

$$\frac{\partial}{\partial z}(\overline{w}+w')(\overline{v}+v')+\frac{\partial}{\partial y}(\overline{v}+v')^2= -\frac{1}{\varrho}\frac{\partial p}{\partial y} \qquad (3)$$

$$\frac{\partial}{\partial z}(\overline{w}+w')(\overline{\vartheta}+\vartheta')+\frac{\partial}{\partial y}(\overline{v}+v')(\overline{\vartheta}+\vartheta')=0 \tag{4}$$

namely the equation of continuity (1), two equations of motion (2) and (3), and an equation expressing the conservation of heat (4). The symbols are listed at the end of the paper.

We assume the motion to be quasistatic, i.e. p equals the pressure in the undisturbed environment. Or $-\frac{1}{\varrho}\frac{\partial p}{\partial z}-g=g\frac{\varrho_0-\varrho}{\varrho}=g\frac{\overline{\vartheta}+\vartheta'}{\Theta}$ where ϱ_0 denotes the density of the undisturbed environment with temperature Θ and where all density differences are assumed to be caused by the excess temperature $\overline{\vartheta}+\vartheta'$ of the jet.

From the foregoing the equations for the mean motion can be found by rearranging and omitting all first powers of the turbulent quantities. Neglecting $-\frac{1}{\varrho}\frac{\partial p}{\partial y}$ with respect to the turbulent frictional forces the following equations are obtained:

$$\frac{\partial\overline{w}}{\partial z}+\frac{\partial\overline{v}}{\partial y}=0 \tag{1a}$$

$$\frac{\partial}{\partial z}(\overline{w})^2+\frac{\partial}{\partial y}(\overline{w}\,\overline{v})=$$

$$=g\frac{\vartheta}{\Theta}-\frac{\partial}{\partial z}\overline{(w')^2}-\frac{\partial}{\partial z}\overline{(w'v')} \tag{2a}$$

$$\frac{\partial}{\partial z}(\overline{w}\,\overline{v})+\frac{\partial}{\partial y}(\overline{v})^2=-\frac{\partial}{\partial z}\overline{(w'v')}-\frac{\partial}{\partial y}\overline{(v')^2} \tag{3a}$$

$$\frac{\partial}{\partial z}(\overline{w}\,\overline{\vartheta})+\frac{\partial}{\partial y}(\overline{v}\,\overline{\vartheta})=-\frac{\partial}{\partial z}\overline{(w'\vartheta')}-\frac{\partial}{\partial y}\overline{(v'\vartheta')} \tag{4a}$$

Now the following assumptions are introduced:

1. At $y=0$, $\overline{w}=w_m$, $\overline{v}=0$, $\overline{\vartheta}=\vartheta_m$, w_m and ϑ_m being functions of z only.
2. At $y=\infty$, $\overline{w}=\overline{v}=\overline{\vartheta}=0$.
3. $\overline{w}$ as well as $\overline{\vartheta}$ show similar distributions in every level z:

   $\overline{w}=w_m f(\eta)$, $\overline{\vartheta}=\vartheta_m g(\eta)$ with $\eta=\frac{y}{R}$, R being a measure for the "width" of the jet and a function of z only.
4. $\overline{(w')^2}$, $\overline{(w'v')}$ and $\overline{(v')^2}$ are functions of η and proportional to w_m^2, $\overline{w'\vartheta'}$ and $\overline{v'\vartheta'}$ are functions of η and proportional to $w_m\vartheta_m$.
5. For $y\to\infty$ the turbulent values become $=0$ together with the mean values.

Next $\overline{v}$ is eliminated from the equations by integrating the equation of continuity:

$$\overline{v}=-\int\frac{\partial}{\partial z}\overline{w}\,dy=-\int R\frac{dw_m}{dz}f(\eta)\,d\eta+$$

$$+\int w_m\frac{df(\eta)}{d\eta}\eta\frac{dR}{dz}dy=R\frac{dw_m}{dz}\varphi(\eta)+$$

$$+\frac{dR}{dz}w_m\psi(\eta)$$

The additional integration constant must be $=0$ as $\overline{v}=0$ for $y=0$ as well as for $y=\infty$, independent of $R(z)$ and $w_m(z)$ (assumptions 1 and 2).

Introducing the expression for $\overline{v}$ into (2a)—(4a) and integrating over y from $-\infty$ to $+\infty$ one obtains:

$$A\frac{d}{dz}[Rw_m^2]=B\frac{g}{\Theta}\vartheta_m R \tag{2b}$$

$$C\frac{d}{dz}\left[R^2\frac{dw_m}{dz}w_m\right]+D\frac{d}{dz}\left[R\frac{dR}{dz}w_m^2\right]=$$

$$=E\frac{d}{dz}[Rw_m^2] \tag{3b}$$

$$F\frac{d}{dz}[Rw_m\vartheta_m]=0 \tag{4b}$$

where the constants A—F are found from the integration of $f(\eta)$ etc. over η from $-\infty$ to $+\infty$.

Assuming R, w_m and ϑ_m to be powers of z, one immediately finds $R\sim z^1$; $w_m\sim z^0$ and $\vartheta_m\sim z^{-1}$, in agreement with W. Schmidt.

Tollmien's solution is obtained by putting $\vartheta_m=0$, resulting in: $R\sim z^1$, $w_m\sim z^{-1/2}$.

A generalized solution can now be arrived at in the following way:

In equation (3b) the variations of kinetic and frictional energies are interconnected. No other form of energy contributing to the horizontal motion, integration with respect to z gives:

$$CR^2\frac{dw_m}{dz}w_m+DR\frac{dR}{dz}w_m^2=ERw_m^2 \tag{3c}$$

From (4b) and (2b) one derives:

$$A\frac{d}{dz}(Rw_m^2)=\frac{B}{F}\frac{g}{\Theta}\frac{Q}{w_m} \qquad (2c)$$

where

$$Q=FRw_m\vartheta_m \qquad (4c)$$

Integration of equation (2c) with respect to z results in:

$$w_m^3=\frac{1}{R^{3/2}}\left\{\int\frac{3BgQ}{2AF\Theta}\sqrt{R}\,dz+K\right\} \qquad (5)$$

where K is a constant.
Introducing this result into (3c) the following differential equation is obtained for R:

$$qR'+r(R')^2+sRR''=t$$

q, r, s and t being functions of the original constants A, B, C, D, E, F and Q, and a prime denoting a differentiation to z. Finally the following solution is obtained for R':

$$q(R')+r(R')^2-t=C_0R^{-2\frac{r}{s}}\,e^{h(R')} \qquad (6)$$

with $h(R')=\int\frac{qdR'}{r(R')^2+qR'-t}$ and C_0 an integration constant. A special solution is arrived at by putting $C_0=0$:

$$R'=\text{const} \quad \text{or} \quad R=\frac{R_1}{z_1}z,$$

any additional integration constant resulting in another position on the z-axis of the origin only.

R_1 represents the value of R in a reference level, the origin lying at a distance z_1 upstream.

Introduction of $R=\frac{R_1}{z_1}z$ into (2c) and (4c) then leads to the general solution:

$$w_m=\left[\frac{Bgz_1}{AF\Theta R_1}Q+\frac{K}{R_1^{3/2}}\left(\frac{z}{z_1}\right)^{-3/2}\right]^{1/3} \qquad (7)$$

$$\vartheta_m=\frac{Qz_1}{FR_1z}\left[\frac{Bgz_1}{AF\Theta R_1}Q+\frac{K}{R_1^{3/2}}\left(\frac{z}{z_1}\right)^{-3/2}\right]^{-1/3} \qquad (8)$$

Obviously K is connected with the extra velocity with which the jet may be forced to emerge from the nozzle. It is immediately evident that for $K=0$, i.e. for the case of a jet rising exclusively due to buoyancy, W. Schmidt's solution: $R\sim z^1$, $w_m\sim z^0$ and $\vartheta_m\sim z^{-1}$ results, whereas for the case of a cold jet, i.e. a jet where no buoyancy forces play a role, $Q=0$ and $R\sim z^1$, $w_m\sim z^{-1/2}$, which is Tollmien's solution.

The general expressions (7) and (8) pertain to a general two-dimensional heated jet that is forced to leave an infinitely long aperture. It may be stressed that it follows from (7) that the resulting velocity cannot be obtained by simple addition of a Tollmien-velocity and a Schmidt-velocity, but that the third powers of the velocities are additive.

3. Axially symmetric jet

The same reasoning holds for the axially symmetric jet. It leads to the following general expressions for w_m and ϑ_m:

$$w_m=\left[\frac{3}{4}\frac{Bgz_1^2}{AF\Theta R_1^2}Qz^{-1}+\frac{K}{R_1^3}\left(\frac{z}{z_1}\right)^{-3}\right]^{1/3} \qquad (9)$$

$$\vartheta_m=\frac{Qz_1^2}{FR_1^2z^2}\left[\frac{3}{4}\frac{Bgz_1^2}{AF\Theta R_1^2}Qz^{-1}+\frac{K}{R_1^3}\left(\frac{z}{z_1}\right)^{-3}\right]^{-1/3} \qquad (10)$$

where the constants have the same meaning as before. We have again a generalization containing Tollmien's and W. Schmidt's solutions:

$R\sim z^1$, $w_m\sim z^{-1}$ and $R\sim z^1$, $w_m\sim z^{-1.3}$ and $\vartheta_m\sim z^{-5/3}$, respectively.

Equations (9) and (10) were given for the first time by Priestley and Ball (1955). A complete derivation along the lines of the preceding section has been given elsewhere (Schmidt 1957).

4. The effect of Q

In both cases the quantity Q plays an important role. It is a measure for the source strength. For the two-dimensional flow Q represents the source strength per unit length of the orifice; in the axially symmetric case the total source strength equals $2\pi Q$.

It follows from the equations (7) to (10) that in the case of heated jets rising according to W. Schmidt's laws the vertical velocity increases according to $Q^{1/3}$, the temperature

excess according to $Q^{2/3}$, or an increase of the total output by a factor 8 doubles the vertical velocity, whereas ϑ_m increases by a factor 4. This result is of importance for the theory of the diffusion of stack gases in the atmosphere where the rise of the plume due to buoyancy forces is quite essential. The method used here to derive the general expressions for w_m and ϑ_m shows, that it is not necessary to introduce any assumptions on the nature of $f(\eta)$ and $g(\eta)$, as was done by Priestley and Ball. The constants A, B, and F depend on the functions of η that have been introduced and they remain arbitrary, therefore, in the present developments.

More general solutions may be obtained by starting from the complete equation (6), i.e. without putting $C_0 = 0$. All experiments seem to indicate, however, that $R' =$ cst is in close agreement with reality except when the environment is in stable equilibrium in which case the jet spreads rapidly above a certain level (Morton c.s. 1956).

5. The velocity profile

It may be worth while to make some remarks with respect to the functions $f(\eta)$ and $g(\eta)$ pertaining to the distribution of velocity and temperature over a cross section perpendicular to the axis of the jet.

Tollmien and W. Schmidt tried to determine both functions from the equations of motion by expanding the stream function into an infinite series. To obtain $f(\eta)$ they introduced as boundary conditions $\bar{v} = 0$ $\bar{u} =$ finite for $\eta = 0$ and $\bar{u} = 0$ for $\eta = \eta_r$, the "boundary" of the jet for the two-dimensional case and $\bar{v} = 0$, $\bar{u} =$ finite for $\eta = 0$ for the axially symmetric jet.

The method, though seemingly exact, has disadvantages. In the first place the solution can be applied to values of $\eta > \eta_r$, but it seems to have no physical significance there. Secondly it does not seem quite correct that under stationary conditions, only attained over an infinitely long time after the motion started, the jet should have a definite "boundary" and should not show any effect up to an infinite distance from the axis, at least theoretically. It is for these reasons that some authors adopted an error function for $f(\eta)$ and $g(\eta)$, e.g. Rouse, Yih and Humphreys (1952), and Priestley and Ball (1955). Though the error function may represent a

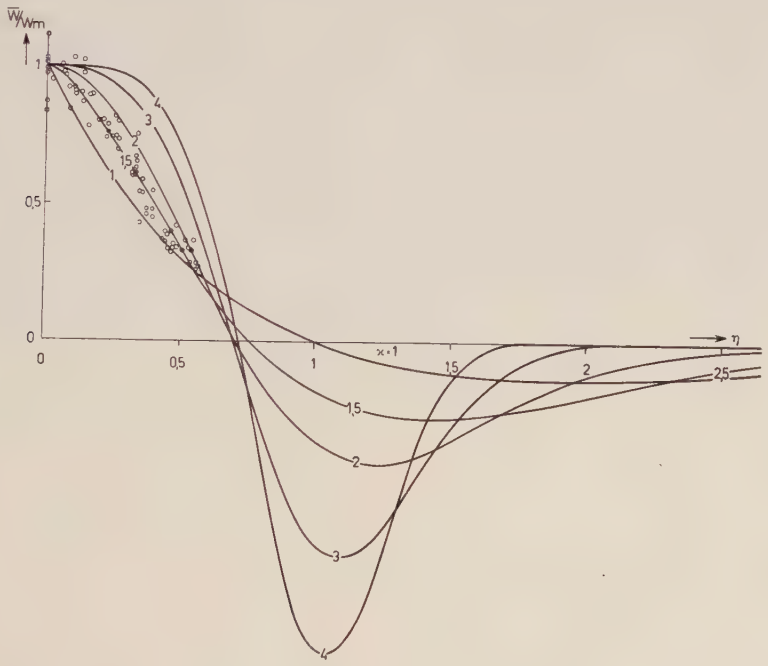


Fig. 1. The distribution of w in a two-dimensional jet as a function of η and $\varkappa$. The small circles pertain to the experimental results of Rouse c.s.

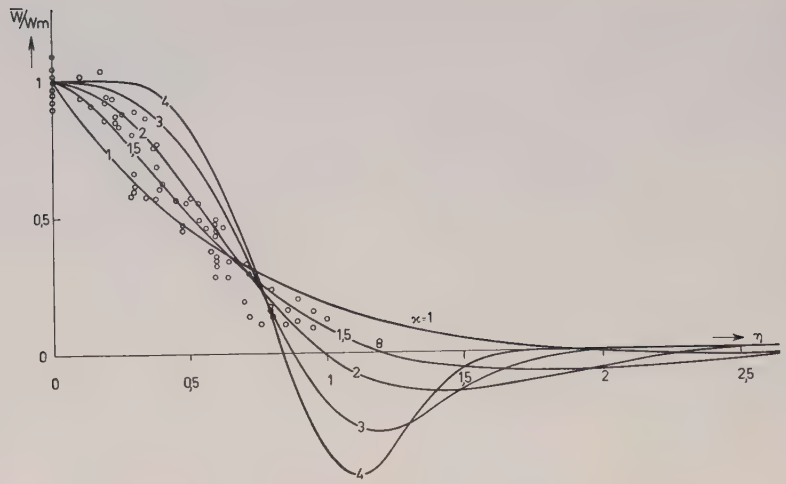


Fig. 2. The distribution of $\overline{w}$ in an axially symmetric jet as a function of η and $\varkappa$. The small circles pertain to the experimental results of Rouse c.s.

reasonable approximation of the temperature profile, it cannot be used to describe the velocity profile.

For, introducing $f(\eta) = \exp\{-\gamma^2/R^2\}$ in the case of the two-dimensional jet, one finds from equation (1a):

$$\overline{v}_\infty - \overline{v}_0 = -\frac{\partial}{\partial z}\int_0^\infty \overline{w}\,dy = -\frac{\partial}{\partial z}\left[\frac{\sqrt{\pi}}{2} R w_m\right] \quad (11)$$

This result shows that $\overline{v}$ cannot be $= 0$ for $\gamma = 0$ as well as for $\gamma = \infty$ except in case Rw_m is independent of z, which is certainly not true for free jets. However, $\overline{v}(\gamma = 0) = 0$ is obviously necessary from a physical point of view as otherwise there would exist a discontinuity for $\gamma = 0$, the plane of symmetry. On the other hand a finite horizontal velocity at $\gamma = \infty$ would involve an infinitely large total kinetic energy per slice of unit thickness and per unit length of the aperture:

$$\int_0^\infty \frac{1}{2}\varrho\,(\overline{w}^2 + \overline{v}^2)\,dy.$$

Moreover, all experimental investigations on the propagation of jets have been performed in a finite room with fixed walls at a distance $\gamma > R$ from the symmetry plane. At these walls $\overline{v}$ must be zero of necessity. The same reasoning holds for convective currents in the atmosphere.

It follows then from (11), that $\int_0^\infty \overline{w}\,dy$ must be constant. From the properties of Γ-function it can be shown easily that

$$\overline{w} = w_m(z)\{1 - \varkappa\eta^\varkappa\}\,e^{-\eta^\varkappa}, \quad (12)$$

where $\varkappa$ is a constant is an example of a function satisfying this condition.

Figure 1 shows the velocity profiles for some values of $\varkappa$. It appears that a downward countercurrent exists for $\eta > \varkappa^{-1/\varkappa}$. This countercurrent has long been recognized in the atmosphere where it occurs around cumuliform clouds. It has not been found in the various experiments on the propagation of jets. It is possible, however, that measurements of $\overline{w}$ have not been performed for sufficiently large values of η.

In figure 1 the points corresponding to the measurements of Rouse c.s. have also been entered. It appears that the $\overline{w}$-curve proposed here is well in accordance with the experimental data for $\varkappa = 1.5$.

The same reasoning applies to the axially symmetric jet leading to the velocity distribution:

$$\overline{w} = w_m(z)\left\{1 - \frac{\varkappa}{2}\eta^\varkappa\right\}e^{-\eta^\varkappa} \quad (13)$$

Figure 2 shows the various velocity distributions and again the experimental results

of Rouse c.s. Theory and experiment are well in accordance for $\varkappa = 1.8$.

6. Final Remarks

It should be kept in mind that the foregoing developments are based on too restrictive assumptions.

No account has been taken of the fact that the similarity of velocity and temperature profiles only comes into being at distances of about ten times the diameter of the aperture from the orifice (KUETHE, 1935).

Moreover, it has been found that the flow in a free jet is a very complicated hydrodynamical phenomenon. For instance, at some distance from the symmetry-axis turbulence appears to be intermittent whereas at large distances from the axis the motion is almost a non-turbulent one (PAI, 1954). Slight improvements in Tollmien's and W. Schmidt's results have been obtained or may be obtained by considering these facts. They have of necessity been ignored here. Nevertheless the present results may be helpful in the interpretation of future experiments as well as in understanding some convective phenomena in the atmosphere.

The author wants to thank Prof. W. Bleeker and Mr. W. J. A. Kuipers for having contributed valuable remarks.

List of symbols

x, y, z	rectangular coordinates, z pointing upwards and y perpendicular to the plane of symmetry in the case of a two-dimensional jet.
$\overline{w}$	mean velocity in the z-direction.
$\overline{v}$	mean velocity in the y-direction.
w', v'	turbulent velocity components in the z- and y-directions.
$\overline{\vartheta}$	mean temperature excess of the jet.
ϑ'	turbulent fluctuation superimposed on $\overline{\vartheta}$.
w_m	mean vertical velocity in the symmetry plane (c.q. axis) of the jet.
ϑ_m	mean temperature excess in the symmetry plane (c.q. axis) of the jet.
Θ	temperature of the environment.
p	pressure.
g	acceleration of gravity.
ϱ	density.
ϱ_0	density of the environment.
R	measure for the width of the jet.
η	y/R.
A—F, q—t,	constants.
Q	measure for the heat output of the jet.

REFERENCES

KUETHE, A. M., 1953: Investigations of the turbulent mixing regions formed by jets; *Journ. Appl. Mech.* **2** p. A 87.

MORTON, B. R., TAYLOR, SIR G., and TURNER, J. S., 1956: Turbulent gravitational convection from maintained and instantaneous sources; *Proc. Roy. Soc., A* **234** p. 1.

PAI, SHIH-I, 1954: *Fluid dynamics of jets;* Van Nostrand Comp., New York, Chapter V.

PRIESTLEY, C. H. B., and BALL, F. K., 1955: Continuous convection from an isolated source of heat; *Qu. J. Roy. Met. Soc.*, **81** p. 144.

ROUSE, H., YIH, C. S., and HUMPHREYS, H. W., 1952: Gravitational Convection from a Boundary Source; *Tellus* **4**, p. 201.

SCHMIDT, F. H., 1957: On the diffusion of stack gases in the atmosphere; *Koninklijk Nederlands Meteorologsch Instituut, Mededelingen en Verhandelingen 68.*

SCHMIDT, W., 1941: Turbulente Ausbreitung eines Stromes erhitzter Luft; *Z.A.M.M.* **21** p. 265 and p. 351.

TOLLMIEN, W., 1926: Berechnung der turbulenten Ausbreitungsvorgänge; *Z.A.M.M.* **4** p. 468.

Electric Charge Separation in Subfreezing Cumuli

By S. TWOMEY, Division of Radiophysics, C.S.I.R.O., Sydney, Australia

(Manuscript received November 22, 1956)

Abstract

It is shown that charges may be separated in large cumuli as a result of negative charging of growing ice particles by collision with supercooled droplets. The magnitude of the effect has been calculated and it has been shown that charges comparable to those involved in thunderstorms can develop, providing the water content and updraught velocity in a cloud are high. The hypotheses lead to conclusions which agree, qualitatively and quantitatively, with the known physical characteristics of thunderclouds.

1. Introduction

The generation of the large electric charges in thunderstorm clouds has been the subject of many theoretical and experimental investigations. Although several theories have been advanced to explain thunderstorm charge-generation processes, most of these theories were expressed only in qualitative form and all seem unsatisfactory when examined quantitatively. In a recent review MASON (1953 a) has outlined and discussed the classicial theories of charge generation and has also described the main physical features of a thunderstorm as deduced from the observations of Byers, Braham, Chapman, Gunn, Kuettner, Malan, Reynolds, Schonland, Simpson, Wilson, Workman, and others. The more important of these features are as follows.

(1) The average electrical moment destroyed in a lightning flash is about 110 coulomb kilometres and the electrical charge involved is 20 to 30 coulombs.

(2) Lightning flashes occur at average intervals of about 20 seconds, so that the current dissipated and the charging current are of the order of 1 amp.

(3) A thundercloud has a bipolar electrical structure, the upper positive charge being centred in the vicinity of the $-30°$ C isotherm, and the lower negative charge centred in the vicinity of the $-8°$ C isotherm. A subsidiary and smaller charge (about 4 coulombs) may also be found below the 0° C isotherm.

(4) The first lightning flash appears within 12 to 20 minutes of the appearance of precipitation particles large enough to be detected by radar.

(5) It is probably necessary for the cloud base to be warmer than 0° C and for the top of the radar echo from the cloud to approach the $-30°$ C level.

(6) An incipient thundercloud is characterized throughout its volume by updraughts averaging 7 m sec^{-1} but values of 15 m sec^{-1} have been frequently found.

(7) The average duration of precipitation and electrical activity is about 30 minutes.

In a subsequent paper MASON (1953 b) outlined a theory for charge generation associated with graupel formation, and showed that charges of the order of 500 coulombs could be separated during the later stages of growth of large ice particles, when the particle radii approached 1 cm. Although the mechanism described may undoubtedly give rise to large charges it does not seem sufficient for the

following reasons to account for most occurrences of thunderstorms.

(1) As Mason himself points out, hailstones of radius 1 cm or more are probably not a normal feature of thunderstorms.

(2) The volume distribution of particles computed by Mason, and their associated charges, must give rise to very high values for precipitation intensity and precipitation current density – of order 2 mm sec^{-1} and 2×10^{-10} amp cm^{-2}, respectively. Both values appear excessive: a precipitation rate of 100 mm per hour is unusual, while CHALMERS and LITTLE (1939) have reported one occurrence of a precipitation current of 7×10^{-12} amp cm^{-2} during a hailstorm as a very unusual event.

In the present paper a different approach has been adopted, based on actual observations of natural cloud element electrification. The charge-separation process postulated is similar to that assumed by Mason, but it will be shown that large charges can be separated without the formation of giant hailstones.

2. The mechanism of electrification

Experiments by FINDEISEN (1940), WORKMAN and REYNOLDS (1948) and several other workers show that the freezing of supercooled water is accompanied by electric charge separation. It now seems certain that the impingement and freezing of supercooled droplets on an ice surface results in a negative charge being acquired by the collecting surface. The balancing positive charge is presumably carried away, either by the surrounding air or by minute droplets. From experiments in natural supercooled cloud, LUEDER (1951) concluded that the charge received by the collecting ice surface increased in direct proportion to the volume of supercooled water collected until it reached a limiting value determined by the surface potential gradient. The charge was given by Lueder as -0.2 coulombs (-6×10^8 e.s.u.) per cm^3 of supercooled water collected, and the limiting surface potential gradient was approximately 700 volts cm^{-1}.

Experimental measurements of cloud element electrification carried out on a mountain top (TWOMEY, 1956) showed that cloud droplets were normally positive, the mean charge being approximately 10^{-9} d^2 e.s.u. (where d was the droplet diameter in microns). When ice crystals were present, however, many of the cloud elements carried negative charges, and it was found that the surface potential gradient, calculated on the assumption that the particles were spherical, appeared to have an upper limit close to 700 volts cm^{-1}. This suggested that the negative charges arose as a result of collisions between ice particles and supercooled cloud droplets and that charging ceased when the surface potential gradient attained a value near 700 V cm^{-1}. It should be noted that, on the basis of a charge of 0.2 coulombs (6×10^8 e.s.u.) per cm^3, the collection of a supercooled droplet of radius 1 micron is more than sufficient to charge a 300 micron particle to the limiting charge determined by the restriction of the surface potential gradient to 700 V cm^{-1}.

3. Assumptions

In order to compute the charge separation which may occur in a cumulus cloud by the mechanism described, the following assumptions have been made.

(1) In an ascending parcel of air, the number of ice crystals which have appeared in the parcel when it has been cooled to a temperature $-T°$ C is the same as the number of freezing nuclei active at that temperature.

(2) The freezing nucleus spectrum is such that a tenfold increase occurs in the number of effective nuclei for each temperature increment of $-4°$ C. This assumption gives a spectrum close to experimentally observed values (BIGG, private communication). Hence if the lapse rate in a cloud is $\gamma°$ C cm^{-1}, the number of nuclei active at or below a height z above the 0° C isotherm is

$$N(z) = A e^{\alpha Z} \qquad (1)$$

where $\alpha = 0.25\ \gamma \ln 10$. The number A is a function of the concentration of freezing nuclei. On the average about 1 nucleus per litre is found at $-20°$ C; under such conditions A has the value 10^{-8}. However, although the shape of the freezing nucleus spectrum does not vary greatly, concentrations often vary enormously from day to day. Hence A may take values ranging from about 10^{-7} to 10^{-9},

and occasionally may be even as low as 10^{-10} or as high as 10^{-6}.

(3) The vapour pressure in that part of the cloud between the 0° C isotherm and the level at which complete glaciation of the cloud occurs (i.e. above which the liquid phase is absent) is the equilibrium vapour pressure above water. The supersaturation with respect to ice at level z, expressed in degrees C, is then approximately $T/10$ at temperature $-T°$ C giving for the supersaturation

$$S = \gamma \frac{Z}{10} °C \qquad (2)$$

(4) A simplified equation for the growth of spherical ice particles by sublimation has been derived from that given by SQUIRES (1952, p. 64) for the growth of cloud droplets by condensation. The equation used is as follows:

$$r\frac{dr}{dt} = 4 \times 10^{-8} S \qquad (3)$$

(r = radius of spherical ice particle in cm, t = time in seconds, S = supersaturation in degrees C with respect to ice).

(5) The particles are assumed to obey Stokes' Law, so that the free fall velocity is

$$u = kr^2 \text{ cm sec}^{-1} \qquad (4)$$

where k is approximately 1.2×10^6 cm^{-1} sec^{-1}. Temperature variations of viscosity and density have been neglected.

(6) A uniform updraught of V cm sec^{-1} and a uniform lapse rate $\gamma = 5 \times 10^{-5}$ ° C cm^{-1} is taken to exist throughout the cloud.

(7) The collision of a supercooled droplet of volume v cm^3 with an ice particle is capable of imparting a charge of $6\ v \times 10^8$ e.s.u. to the particle, but the charge which can be imparted to a particle of radius r has a maximum value of $7/3\ r^2$ e.s.u., this being the charge which produces a surface potential gradient of 7/3 e.s.u. (700 V cm^{-1}).

(8) Charging ceases when no supercooled droplets are available or when the ice particle melts or acquires a covering of unfrozen water.

Although the above assumptions represent a much simpler model than is found in nature, they do not seem likely to introduce any gross errors in the final deductions.

4. The growth of ice particles

Ice particles grow in cloud by two distinct mechanisms: firstly by sublimation (i.e. diffusion of water vapour from the supersaturated environment) and secondly, by collection of water droplets and ice particles. From equation (2) the rate of increase of volume by sublimation is seen to be given by

$$\left(\frac{dv}{dt}\right)_s = 16\pi \times 10^{-8} Sr.$$

Growth by collection of smaller ice particles and water droplets is given approximately by

$$\left(\frac{dv}{dt}\right) = \pi k E w_c r^4.$$

(w_c = water content of cloud, including liquid and solid phases, and
E = collection efficiency)

Hence

$$\frac{\left(\frac{dv}{dt}\right)_c}{\left(\frac{dv}{dt}\right)_s} = \frac{kEw_c r^3}{16 \times 10^{-8} S}.$$

Substituting values for typical cloud conditions with $w_c = 1$ g m^{-3} and $S = 2°$ C this ratio is found to become unity when the radius approaches 100 μ. Smaller particles grow mainly by sublimation, while larger particles grow mainly by accretion.

If accretion is neglected, the equations of motion for an ice particle above the 0° C isotherm can readily be solved. By combining equations (2), (3) and (4), one obtains

$$\frac{d^2z}{dt^2} = -\omega^2 z$$

where $\omega = (8k \times 10^{-9}\gamma)^{1/2} \approx 7 \times 10^{-4}$ sec^{-1}. The solution of this equation is

$$\left.\begin{aligned} z &= Z \sin(\omega t + \varphi) \\ \text{from which} \\ r &= k^{-1/2}\{V - \omega Z \cos(\omega t + \varphi)\}^{1/2} \end{aligned}\right\}. \qquad (5)$$

If the ice particle came into existence at a height ζ at time $t = 0$, the constants become

$$\varphi = \tan^{-1} \frac{\omega\zeta}{V}$$

$$Z = \zeta \operatorname{cosec} \varphi.$$

(Here, and everywhere in the subsequent discussion, height is measured from the 0° C isotherm.) Equation (5) describes the trajectory and time dependence of radius for an ice particle growing by diffusion alone; it is not, of course, applicable for negative values of z or ζ.

The probability of a particle undergoing a collision with a cloud droplet between t and $t + \Delta t$ is readily found to be

$$p(t)\,\Delta t = \pi k E N r^4\,\Delta t$$

where N is the number of cloud droplets per unit volume and E the corresponding collection efficiency. Hence the probability of a particle originating at time $t = 0$ undergoing at least one collision within a time t is given by

$$P(t) = 1 - e^{-\pi k \int_0^t E N r^4 dt} \qquad (6)$$

The quantity $P(t)$ represents that fraction of the total number of particles originating at time $t = 0$ which have undergone one or more collisions at time t. The exponent in the expression for P(t) is a function of the height at which the particle originated and also of the supersaturation and other cloud parameters. However when $P(t)$ was evaluated for several sets of values of the various parameters it was found that the time τ which gave $P(t) = 0.5$ did not vary greatly, ranging from about 50 to 70 seconds; in this time the particle grew to about 30 microns radius. The calculations were made assuming growth from zero radius at $t = 0$.

5. The collection of charge by growing ice particles

As indicated earlier, the assumptions made concerning the charging mechanism imply that the collection of a single cloud droplet is capable of imparting the maximum charge to all but the largest ice particles. Hence when a growing particle experiences the first collision with a cloud droplet (which occurs in the average some 60 seconds after its formation) it will acquire the maximum charge, which is a function of particle radius. Thereafter frequent collisions will occur, which are potentially capable of separating relatively large charges; the actual charge imparted, however, is limmited by the restriction that the charge carried by a particle of radius r shall not exceed $7/3\ r^2$ e.s.u. Consequently once a particle has undergone its first collision the charge carried will be determined by the particle radius. Hence for $r > 30\ \mu$, the charge is taken as

$$q \approx -2.33\, r^2 \text{ e.s.u.} \qquad (7\,a)$$

and

$$\frac{dq}{dt} = -4.66 \times r \frac{dr}{dt}. \qquad (7\,b)$$

If a rising parcel of cloud air of unit volume at level z is considered, $A\,e^{\alpha z}$ ice crystals have formed during the ascent from the 0° C isotherm; the particles originating at lower levels have grown and fallen from the parcel, but have been replaced by an approximately equal number falling into the parcel from above. Thus there are approximately A $e^{\alpha z}$ ice particles per unit volume at level z and those particles which have been in existence for a time τ or more have undergone several collisions with supercooled droplets and therefore carry charges the magnitude of which are given by equation (7a). The number of such particles is clearly the total number of particles formed when the rising air parcel was at a height $z - \tau V$. The charging rate for such a particle is given by equation (7b). Combining these results with equations (2) and (3), one finds that the total charging rate per unit volume of ascending cloud air at level z is given by

$$\frac{dQ}{dt} = Ae^{\alpha(z-\tau V)} \times 4.66\,(4 \times 10^{-9}\gamma)\,z$$
$$\approx 2 \times 10^{-8} \gamma A z\, e^{\alpha(z-\tau V)}.$$

But the parcel is ascending at velocity V, hence

$$Q = 2 \times 10^{-8} \frac{\gamma A e^{-\alpha\tau V}}{V} \int_{\tau V}^{z} z e^{\alpha z} dz$$

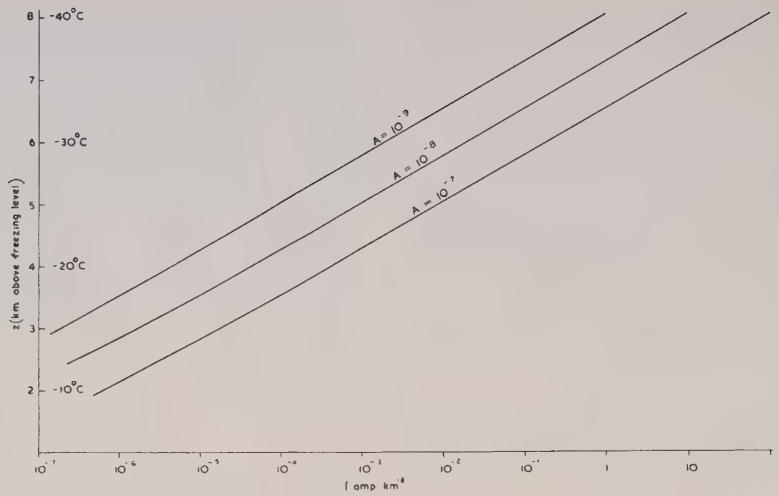


Fig. 1. Graph showing the variation of the function I (z) with height (ż). I(z) represents the current flowing at or below the glaciation level z_M; for $z < z_M$ the current is constant and equal to $I(z_M)$.

and

$$Q(z) \approx 2 \times 10^{-8} \frac{\gamma A}{V\alpha^2} (\alpha z e^{\alpha z} - e^{\alpha z})$$

providing $z \gg \tau V$.

Therefore the current per unit area being carried by the cloud air at level z is found to be

$$i(z) = 2 \times 10^{-8} \gamma A \alpha^{-2} (\alpha z e^{\alpha z} - e^{\alpha z}) \text{ e.s.u. cm}^{-2}. \quad (8)$$

This function converted to amps km^{-2} has been plotted in fig. 1. If at some level z_M charging of the ice particles ceases, the charged ice particles rapidly fall out and the ascending parcel of air is left with a positive charge which is carried aloft and is equivalent to a current $i(z_M)$. It is clear from equation (8) that this current depends critically on z_M, and accordingly it is necessary to determine the value of z_M from the initial assumptions.

6. Evulation of the glaciation level

On the basis of the charging mechanism which is being considered, charge separation in a rising parcel of cloud air will cease when liquid droplets are no longer present, i.e. when complete glaciation of the cloud occurs. Supercooled water droplets have been observed in cumulus clouds at temperatures as low as $-40°$ C, while complete glaciation has occasionally been reported at temperatures as high as $-15°$ C. It may be said that the temperature of complete glaciation ranges from $-15°$ C to $-40°$ C, and is most frequently around $-30°$ C. However, it does not seem desirable to use these experimental data to determine what value of z is to be substituted in equation (8) to obtain the maximum current density; it is preferable to calculate this level on the basis of the assumptions which led to equation (8). This approach has a twofold advantage in that it permits the evaluation to be completed in a self-contained form and also in that the results of a mathematical evaluation of the glaciation level can be compared with the results of observations, thus providing an independent test of the validity of the assumptions made.

In order to determine the glaciation level, it is necessary to compute the total mass of ice per unit volume in a rising parcel of cloud air; complete glaciation is taken to occur at the level at which the mass of ice per unit volume is equal to the mass of available water per unit volume.

Substitution of $dm = 4\pi r^2 dr$

and $dz = V dt$

in equations (2) and (3) gives, for an ice

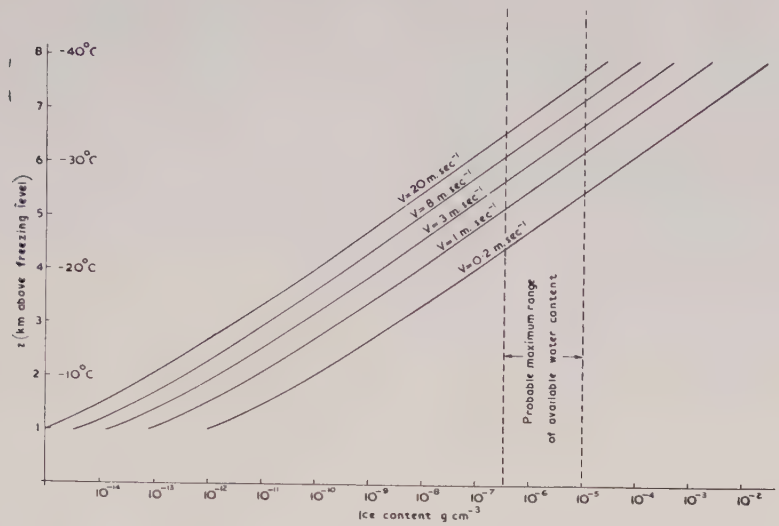


Fig. 2. Graph showing variation of computed ice content with height (z) for values of updraught velocity V from 0.2 m sec^{-1} to 20 m sec^{-1}, and A = 10^{-8} (i.e. 1 freezing nucleus per litre active at — 20° C).

particle originating at a height ζ and growing from zero radius,

$$dm = (16\pi\, 10^{-9}\gamma\, V^{-1})\, rz\, dz$$

Hence

$$m = (16\pi \times 10^{-9}\gamma V^{-1}) \int_{\zeta}^{z} rz\, dz.$$

Also

$$r dr = 4 \times 10^{-9}\gamma\, V^{-1} z\, dz,$$

whence

$$r^2 = 4 \times 10^{-9}\gamma\, V^{-1}(z^2 - \zeta^2)$$

and

$$\int_{\zeta}^{z} rz dz = (4 \times 10^{-9}\gamma V^{-1})^{1/2} \int_{\zeta}^{z} z(z^2 - \zeta^2)^{1/2} dz$$

$$= \frac{1}{3}(4 \times 10^{-9}\gamma V^{-1})^{1/2}(z^2 - \zeta^2)^{3/2}.$$

Thus the mass at a height z of a particle which originated with zero radius at level ζ is given by

$$m = \frac{32\pi}{3}\left(\frac{10^{-9}\gamma}{V}\right)^{3/2}(z^2 - \zeta^2)^{3/2}.$$

But the number of particles originating per unit volume between ζ and $\zeta + d\zeta$ is $A\alpha e^{\alpha\zeta} d\zeta$, from which one finds the total mass per unit volume at level z to be

$$\mu(z) = \frac{32\pi}{3}\left(\frac{10^{-9}\gamma}{V}\right)^{3/2} A\alpha \int_{0}^{z} e^{\alpha\zeta}(z^2 - \zeta^2)^{3/2} d\zeta. \quad (9)$$

The integral here is to be evaluated with z held constant and ζ as the variable of integration. If this integral is denoted by $I(z)$,

$$I(z) = \int_{0}^{z} e^{\alpha\zeta}(z^2 - \zeta^2)^{3/2} d\zeta$$

$$= \frac{z^4}{2} \int_{0}^{1} e^{\alpha z x^{1/2}} x^{-1/2}(1 - x)^{3/2} dx$$

$$= \frac{z^4}{2} \sum_{r=0}^{r=\infty} \left\{ \frac{(\alpha z)^r}{r!} \int_{0}^{1} x^{1/2(r-1)}(1 - x)^{3/2} dx \right\}$$

$$= \frac{z^4}{2} \sum_{r=0}^{r=\infty} \frac{(\alpha z)^r}{r!} B\left(\frac{r+1}{2}, \frac{5}{2}\right),$$

where $x = \dfrac{\zeta^2}{z^2}$ and $B(p, q)$ is the complete Beta-function $\int_{0}^{1} y^{p-1}(1 - y)^{q-1} dy$. When this result is substituted in equation (9), one obtains

$$\mu(z) = \frac{16\pi}{3}\left(\frac{10^{-9}\gamma}{V}\right)^{3/2} A\alpha z^4 \sum_{r=0}^{r=\infty} \frac{(\alpha z)^r}{r!} B\left(\frac{r+1}{2}, \frac{5}{2}\right). \quad (10)$$

Using the last result, the mass of ice per unit volume as a function of height has been cal-

culated; the variation of computed mass with height is shown in fig. 2. According to equation (10), the mass per unit volume of ice is in direct proportion to the concentration of freezing nuclei. Hence if the freezing nucleus spectrum is such that there are 10 nuclei per litre active at $-20°$ C, the masses given in fig. 2 should be multiplied by 10, while for 10^{-1} nuclei per litre active at $-20°$ C the masses should be divided by 10.

Under adiabatic conditions most of the water vapour in an ascending parcel of air condenses to liquid water in the first few km above the condensation level. Hence at higher levels in a cloud the amount of available water is determined mainly by the water vapour content (mixing ratio) of the air at cloud base. This in turn is determined primarily by the cloud base temperature; in warm moist air up to 12 g m^{-3} may be available and in cold dry air only 2 g m^{-3}. It has however been found that the measured liquid water content in cumulus clouds is only a fraction of the adiabatic value. The observations of WARNER (1955) showed that at the higher levels in cumuli about 3 km deep the liquid water content was about one-fifth of the adiabatic value. If this holds in the case of large cumuli, the available water in the upper parts of such clouds might better be taken as ranging from 0.4 g m^{-3} to 2.5 g m^{-3}.

The glaciation level has been computed for several values of V, A and available water content w_c, and the results are set out in table 1. A lapse rate of 5×10^{-5}° C cm^{-1} has been assumed here and in all other computations involving lapse rate.

From table 1 it is seen that the computed glaciation levels agree well with observational results; the table shows that the computed glaciation level may range from 3.75 km ($-18.75°$ C), for small updraught, low water content and high freezing nucleus concentration, to 8.4 km ($-42°$ C) for large updraught and water content and low freezing nucleus concentrations.

From equations (8) and (10) it follows that both current and ice content at any height are directly proportional to A (which determines freezing nucleus concentration), which implies that an increase in A results in an increase in current at any given height, but also results in a lowering of the glaciation level. The latter gives the value z_M which must be used to determine the maximum current i_M from equation (8), so that the maximum current does not vary appreciably with A, but the level at which it is attained is lowered significantly as A increases. For example, if the updraught and water content are taken as 3 m sec^{-1} and 2 g m^{-3} respectively, the glaciation level increases from 5.55 km for $A = 10^{-7}$ to 7.0 km for $A = 10^{-9}$, but the resulting change in i_M is insignificant – from 4.5×10^{-2} amp km^{-2} to 5×10^{-2} amp km^{-2}. The variation of liquid water content with height is very slow several km above cloud base and can be ignored. The above considerations imply that maximum charge separation can occur in smaller clouds if the freezing nucleus concentration is high.

Table 1. Computed values for glaciation level (km above 0° C isotherm

(a) $A = 10^{-8}$ cm^{-3} (1 freezing nucleus per litre effective at $-20°$ C)

V (m sec^{-1})	w_c (g m^{-3})				
	0.4	1.0	2.0	4.0	10.0
0.2	4.45	4.75	4.95	5.2	5.45
1.0	5.2	5.5	5.75	5.95	6.25
3.0	5.75	6.05	6.25	6.5	6.75
8.0	6.2	6.5	6.7	6.95	7.25
20.0	6.6	6.95	7.15	6.4	7.7

(b) $A = 10^{-7}$ cm^{-3} (10 freezing nuclei per litre effective at $-20°$ C)

V (m sec^{-1})	w (g m^{-3})				
	0.4	1.0	2.0	4.0	10.0
0.2	3.75	4.05	4.25	4.45	4.75
1.0	4.45	4.8	5.0	5.2	5.5
3.0	5.0	5.3	5.55	5.75	6.05
8.0	5.45	5.8	6.0	6.2	6.5
20.0	5.9	6.2	6.4	6.6	6.95

(c) $A = 10^{-9}$ cm^{-3} (0.1 freezing nucleus per litre effective at $-20°$ C)

V (m sec^{-1})	w (g m^{-3})				
	0.4	1.0	2.0	4.0	10.0
5.2	5.2	5.45	5.7	5.9	6.2
1.0	5.95	6.25	6.45	6.7	7.0
3.0	6.5	6.75	7.0	7.2	7.5
8.0	6.95	7.25	7.45	7.65	7.95
20.0	6.4	7.7	7.9	8.1	8.4

Table 2. Current at glaciation level (amp km^{-2}) for several values of V and w_c

V (m sec^{-1})	w_c (g m^{-3})				
	0.4	1.0	2.0	4.0	10.0
0.2	0.00018	0.00045	0.00085	0.0018	0.004
1.0	0.0018	0.005	0.01	0.018	0.05
3.9	0.01	0.025	0.05	0.10	0.20
8.0	0.04	0.10	0.18	0.40	1.0
20.0	0.13	0.40	0.70	1.7	4.0

In table 2, the maximum current density has been tabulated for several values of updraught velocity V and available water content w_c. It will be noticed that current densities sufficient to provide the charging rate of about 1 amp required in thunderclouds are given by the higher values of updraught velocity and available water content. For example if the available water content is 4.0 g m^{-3}, a current of 1 amp is provided by an updraught of 3 m sec^{-1} extending over 10 km^2, or an updraught of about 15 m sec^{-1} over an area of 1 km^2.

7. Electrical moment separation

On the basis of the theory described, charge separation commences as soon as the ice crystals begin to form in a growing cumulus, but does not become significant until the top of the cloud is 5—6 km above the 0° C isotherm (i.e. until the cloud top temperature is −25° C to −30° C). Thereafter rapid charge separation occurs and within 10—20 minutes several thousands of coulombs may have been separated. Much of this charge, however, may be located in that part of the cloud where the positive charge on the air is largely balanced by the presence of negatively charged ice particles. It is readily seen that settling of the growing ice particles leads to a bipolar charge distribution of the correct sign, and the resulting electrical moment separation can be estimated by considering the falling velocity of the ice particles. In 10 minutes ice particles can grow by sublimation alone to a radius of order 100 μ, so that the falling velocity of the particles as given by Stokes' Law is of order 1.3 metres per second. If for example a total charge of 2,500 coulombs has been generated, one obtains for the moment separation a value of about 3.25 coulomb kilometres per second. Thus the 110 coulomb km destroyed in an average lightning flash is separated in about one half-minute.

8. The charge on precipitation and the distribution of negative charge

If the liquid droplets in the warmer parts of the cloud carry positive charges of the order of 4×10^{-1} a^2 e.s.u. (a = cloud droplet radius in cm), as was found in experimental observations, the negative charge acquired by a growing ice particle in the upper part of the cloud will tend to be neutralized when it settles into the lower parts of the cloud. Once an ice particle has melted or acquired a liquid covering (LUDLAM, 1950), the negative charging process will cease and the falling particle will commence to acquire a positive charge by collection of positive cloud droplets. Of for example an ice particle melts at radius r_0 while carrying the postulated maximum charge of $-2.33\ r_0^2$ e.s.u., and falls through a cloud containing 200 droplets per cm^3 of radius 10 μ, which carry positive charges of order 4×10^{-7} e.s.u., it can readily be shown that the charge originally carried by the ice particle will be neutralized in a fall of about 0.1 km.

It follows from the above considerations that the final charges on precipitation elements depend on the electrification of the cloud droplets near cloud base and bear little relation to the main charge-separation process. In fact either positive or negative charges may be found on precipitation. A more important consequence is that the many thousands of coulombs involved in thunderstorm processes need not be removed from the cloud itself by precipitation: if the negative particles become neutralized by collection of positively charged cloud droplets, the negative charge on the cloud air, which originally was balanced by positive charges carried by the cloud droplets, becomes "exposed" so that a negative charge is acquired by the cloud air near the 0° C isotherm. This negative charge will be carried up with the rising air, hence the negative charge centre should extend up from the 0° C isotherm. MALAN and SCHONLAND (1951 a, 1951 b) reported that the negative charge involved in a lightning flash to earth

was contained in a vertical column up to 6 km deep and that successive strokes tapped higher and higher regions of the negative column. These observations are consistent with the negative charge distribution suggested.

9. Discussion

Before discussing the conclusions to which the development of the initial assumptions have led, it is desirable to refer to the effect of varying the parameters which have been treated as constants. It is found that the most critical assumptions are those involving the growth equations of ice particles and the freezing nucleus spectrum.

The numerical factor 4×10^{-8} in equation (3) is not in fact a true constant. It decreases with decreasing temperature (SQUIRES, loc. cit.), so that the growth rate at higher levels may have been overestimated. However, the departure of the particles from sphericity and the contribution of accretion to the growth rate would tend to have the opposite effect, and these factors also have not been taken into account. It is probable that the best test of the assumptions made concerning growth rate is obtained by comparing the computed glaciation levels with the results of observations. The close agreement found suggests that the growth equation adopted was a fair approximation.

The shape of the natural freezing nucleus spectrum is not constant, so that the quantity α is also likely to vary in nature. The assumed value of α was 2.88×10^{-5}, which corresponded to a logarithmic freezing nucleus-temperature spectrum such that the total number of effective nuclei increased by a factor of ten for a temperature decrement of 4° C. Experimental observations suggest that the temperature interval so designated may on occasions range from 2° C to 6° C, i.e. α may range from 5.75×10^{-5} to 1.92×10^{-5}. When the effect of varying the value of α is calculated, it is found that the maximum current $i(z_M)$ is almost independent of α, whereas the glaciation level (z_M) varies markedly with α. For example, table 1 shows that for w_c in the range 0.4 – 10.0 g m^{-3} and V in the range 0.2 – 20 m sec^{-1}, $\alpha = 2.88 \times 10^{-5}$ and $A = 10^{-8}$, the computed glaciation level ranged from 4.45 to 7.7 km. If the above extreme values of α are employed, one finds the range of z_M to be 2.6 – 4.2 km for $\alpha = 5.75 \times 10^{-5}$ and 6.2 – 10.8 km for $\alpha = 1.92 \times 10^{-5}$. It is seen that the computed values for the glaciation level are still feasible, while the maximum currents are not appreciably altered. It follows that the conclusions which have been reached are not drastically modified by altering the value of α, within the bounds of its observed variations.

On the basis of the initial assumptions, which were largely based on experimental evidence, a model has been set up for a thunderstorm cell. This has been shown to lead to certain conclusions, of which the more important are repeated below.

(1) In a large cumulus extending well above freezing level, large electric charges can be separated as a result of the collection of supercooled droplets by growing ice particles.

(2) The resulting charge distribution is such that a positive charge is accumulated in the upper part of the cloud, with a lower negative charge extending from just below the 0° C isotherm to several km above.

(3) The upper charge in average conditions will mainly be centred 6 km to 7 km above freezing level (i.e. between the −30° C and the −35° C levels).

(4) The rate of charge separation increases rapidly with increasing height above freezing level but ceases when the glaciation level is reached. The current therefore attains a maximum just below glaciation level.

(5) The maximum current density increases with increasing updraught velocity and water content, the computed values ranging from 0.0018 amp km^{-2} for an updraught of 0.2 metres per second and a water content of 0.4 g m^{-3} to 4 amp km^{-2} for an updraught of 20 metres per second and a water content of 10 g m^{-3}. This dependence on updraught and water content implies that vigorous cumuli, the bases of which extend well below the 0° C isotherm, are the most likely to become thunderclouds.

(6) The level at which current reaches its maximum value depends on the freezing nucleus concentration (assuming a one-to-one correspondence between freezing nuclei and ice crystals). The cloud must penetrate to this level to achieve maximum electrical activity; this critical level becomes lower as freezing

nucleus concentration increases, being lower by about 1.5 km for a hundredfold increase in nucleus concentration. This suggests that in otherwise similar conditions thunderstorms are more likely to occur when the freezing nucleus concentration is high.

(7) The level to which a cloud must penetrate in order to realize its maximum electrical activity is under average conditions, 6—7 km ($-30°$ C to $-35°$ C).

Despite the simplifications made in developing the theory, it is seen that the above conclusions agree well with the main known physical features of a thunderstorm as given by MASON (1953 a) and summarized in the introductory paragraph. The close agreement between the computed glaciation level and the results of observations suggest that the assumptions made were not grossly in error. The suggested dependence of minimum cloud top height for maximum electrical activity on freezing nucleus concentration, and the consequent implication that thunderstorms are more likely when freezing nucleus concentrations are high, may provide a further test of the validity of the hypotheses which have been made.

10. Conclusions

It seems likely that charge separation during the collection of supercooled water droplets by growing ice particles plays an important role in thunderstorm charge generation. It has been shown that current densities of order 1 amp km^{-2} can be produced as a result of the charging of ice particles in this fashion and that the conditions necessary for the generation of large charges are similar to those which have been deduced from the known properties of thunderstorms.

REFERENCES

CHALMERS, J. A., and LITTLE, E. W. R., 1939: Electric charge on soft hail. *Nature*, **143**, p. 244.

FINDEISEN, W., 1940: Über die Entstechung der Gewitterelektrizität. *Met. Zeit.*, **57**, p. 201.

LUDLAM, F. H., 1950: The composition of coagulation-elements in cumulonimbus. *Quart. J. r. meteor. Soc.*, **76**, 52—58.

LUEDER, H., 1951: A new electrical effect during the formation of soft hail (graupel) in natural undercooled mist. Electrification during the formation of soft hail (graupel) as a source of thunderstorm electricity. *Z. angew. Phys.*, **3**, 247—253 and 288—295.

MALAN, D. J., and SCHONLAND, 1951 a: The electrical processes in the intervals between the strokes of a lightning discharge. *Proc. Roy. Soc.* A, **206**, 145—163.

MALAN, D. J., and SCHONLAND, B. F. J., 1951 b: The distribution of electricity in thunderclouds. *Proc. Roy. Soc.* A, **209**, 158—177.

MASON, B. J., 1953a: A critical examination of theories of charge generations in thunderstorms. *Tellus*, **5**, 446—460.

MASON, B. J., 1953b: On the generation of charge associated with graupel formation in thunderstorms. *Quart. J. r. meteor. Soc.*, **79**, 501—509.

SQUIRES, P., 1952: The growth of cloud drops by condensation. I —General characteristics. *Aust. J. Phys.*, **5**, 59—86.

TWOMEY, S.: The electrification of individual cloud droplets. *Tellus*, **8** pp. 443—444.

WARNER, J., 1955: The water content of cumuliform cloud. *Tellus*, **7**, 449—457.

WORKMAN, E. J., and REYNOLDS, S. E., 1948: A suggested mechanism for the generation of charge in thunderstorm electricity. *Phys. Rev.*, **74**, 709.

A new Technique for Counting Ice-Forming Nuclei in Aerosols

By E. K. BIGG, Division of Radiophysics, C.S.I.R.O., Sydney, Australia

(Manuscript received September 9, 1956)

Abstract

An objective method for measuring freezing nucleus concentration with improved accuracy is described and the results of a series of measurements made on the coastline in an on-shore wind near Carnarvon, Western Australia are given. Fluctuations in ice crystal counts existed which were well beyond the probable experimental error and appeared to be independent of weather conditions, except for the occurrence of cirrus cloud. A remote terrestrial source or an extraterrestrial source of nuclei is indicated.

1. Introduction

The use of silver iodide to stimulate rainfall has increased interest in the measurement of naturally-occurring freezing nuclei. The widely varying temperatures at which clouds in the free atmosphere have been observed to contain the ice phase suggest that the nucleus concentration is far from homogeneous. Yet few measurements have been made on a continuous basis over a long period in order to observe these variations in concentration. Some recent exceptions are the experiments described by SMITH, KASSANDER and TWOMEY (1956) in which daily observations were made for about one month. In common with many earlier experiments, the counting technique was not entirely objective so that the real extent of the fluctuations was uncertain.

The present paper describes first a simple and objective method of measuring ice crystal concentrations in an artificially formed supercooled cloud, and then presents the results of a 3 months' series of observations using this method.

SMITH and KASSANDER avoided possible effects of local pollution by making their measurements from an aircraft, while TWOMEY conducted his experiments on the top of 10,000-ft mountain in the Hawaiian Islands. The present measurements were made on the coast at Carnarvon in Western Australia where reliable on-shore winds were known to exist. This site had the further advantages for these experiments of uniform weather and little cloud or rainfall.

2. Measurements of freezing nucleus concentrations

Most freezing nucleus measurements so far reported have been based on visual estimates of the concentration of ice crystals in a strong beam of light. The method relies on the ability of the operator to estimate concentrations consistently, to reproduce conditions such as the density of the cloud and to decide correctly what is an ice crystal and what is not. The latter is made difficult if there are particles in the atmosphere which scintillate, but are not ice, and there may be ice crystals, irregular or partly formed which do not scintillate. CWILONG's (1947) apparatus overcame these difficulties at the expense of reducing the information obtained. In the small expansion chamber which he used was a pool of supercooled water, and its freezing signified that one or more ice crystals had fallen into it.

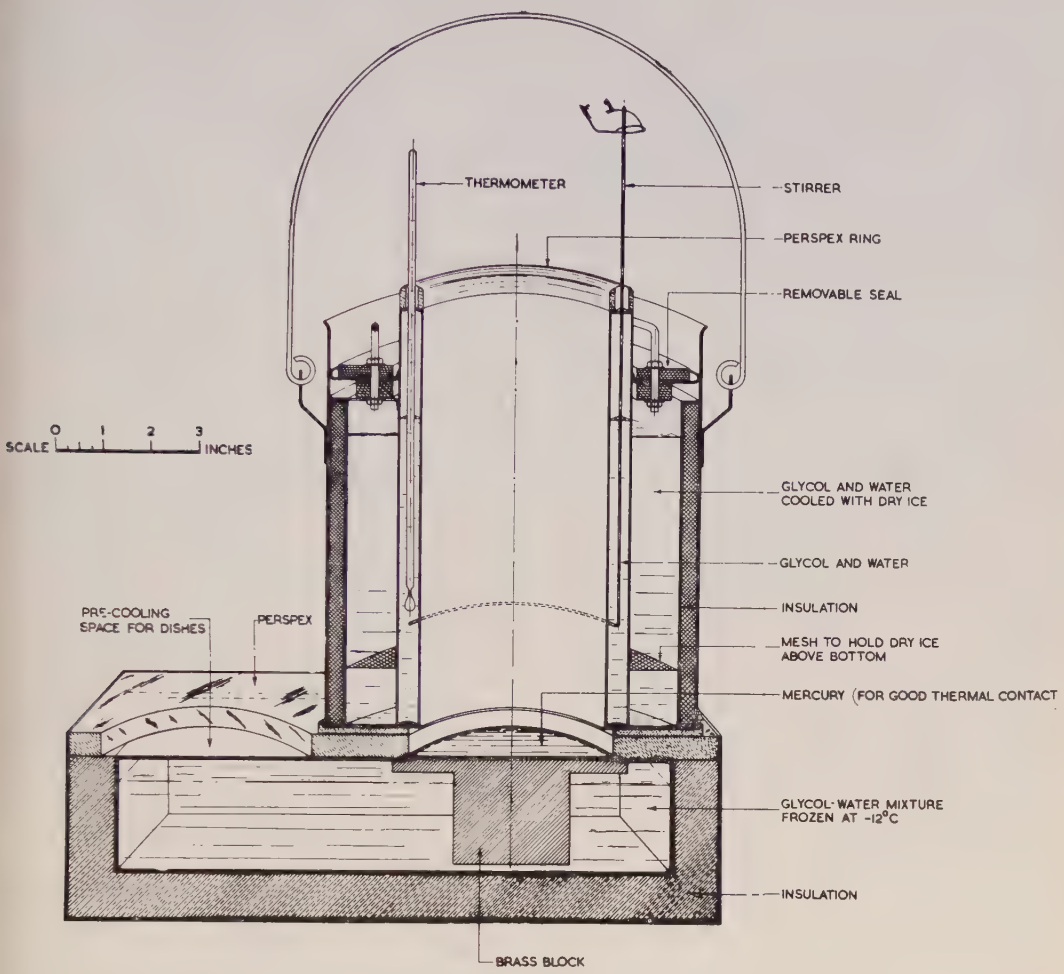


Fig. 1. Cloud chamber and cooling tank.

The information was therefore limited to a threshold temperature for the appearance of ice crystals. The small sampling volume and very short time at minimum temperature were further disadvantages.

The procedure adopted in the present case was a modification of CWILONG's method in which the supercooled water on the bottom of the cold chamber was replaced by a strong aqueous solution. Ice crystals falling into it from the cloud did not freeze the whole of the solution, but grew as individual crystals at a rate controllable over a wide range by varying the concentration of the solution, or the degree of supercooling. In practice one of the simplest and least critical solutions to use is cane sugar in water, as it is readily supercooled to low temperatures. As a typical example, 1.2 kg of sugar added to one litre of water will allow crystals to grow to about 2 mm diameter in 2 minutes when held at $-12°$ C, the melting range being -10 to $-7°$ C. At temperatures warmer than $-15°$ C risk of nucleation in the sugar itself by foreign substances is negligible.

A sketch of the cold chamber and the device for supercooling the sugar solutions is given in Fig. 1. The cold chamber consisted of a hollow cylinder (in which the cloud formed) surrounded by two concentric cylinders partly

filled with ethylene glycol-water mixtures. The outer of these was cooled with dry ice, the inner being designed to prevent cold spots. The temperature in the inner jacket, measured after adequate stirring, is the quantity referred to throughout as the "temperature of the cloud". In fact, the average temperature of the cloud is higher, by about 5° C at −20° C but its minimum temperature closely approximated the recorded value.

The cold chamber itself stood on a rectangular thermally insulating base plate in which were two holes to accommodate petrie dishes (not shown in the figure). The petrie dishes contained sugar solution and made good thermal contact with the heavy metal tank below the base plate containing a melting mixture of glycol and water at −12° C.

In making a measurement of freezing nucleus concentration the sequence of events was as follows: the cold chamber was lifted from an insulated box in which it was kept when not in use and placed on the base plate above a petrie dish. The act of lifting the chamber replaced the air within it, and on being lowered to the base plate, a cloud formed together with a small number of ice crystals, which fell into the supercooled solution. After a fixed interval—in this case 2 minutes—the crystals were counted. Although the cooled volume of air was only 1 litre, larger volumes could be sampled by repeating the experiments many times at the one temperature.

Some precautions were essential. Perhaps the most important was the removal of frost from all surfaces in contact with the cloud. This was ensured by careful wiping with glycerol after each measurement. Another possible source of trouble was spontaneous nucleation in the sugar solution. This could occur if the cold sugar solutions absorbed water from the air, or if the temperature fell too low. It was avoided by frequent replacement of the sugar solution on humid days and by careful regulation of the temperature.

As a typical example of the counts obtained, Fig. 2 shows the measurements made at Carnarvon, Western Australia on 19 January 1956. In the diagram, the points marked by crosses are arithmetic averages over the temperature ranges −20.0° to −21.2° C and −21.7° to −22.5° C. It then becomes possible

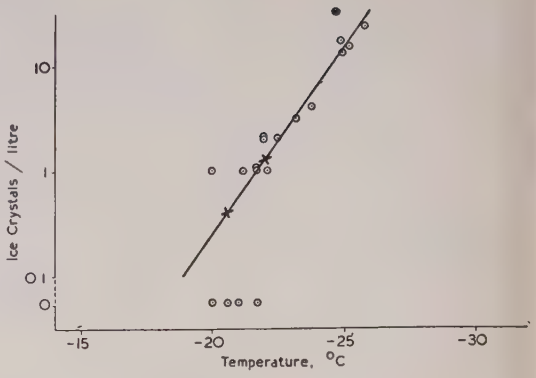


Fig. 2. A typical measurement of freezing nucleus concentration.

to draw a straight line providing a satisfactory representation of the points, the 0.1/l level being reached by a straightforward extrapolation. Occasionally, however, a straight line plot was not acceptable and the number of experiments then had to be increased to provide a reasonable estimate of the temperature for 0.1/l. The number of points necessary for a good determination varied from 15 to 50 depending on the slope of the line (which varied from day to day) and the scatter of the points.

3. Sources of error

The true numbers of nuclei present could differ from the number counted for three main reasons.

(1) Ice crystals may not have grown large enough to fall out of the cloud. It was found that few fell out after 90 seconds although wisps of fog were then still visible and it is considered that the time allowed of 2 minutes was sufficient.

(2) The high probability of obtaining a count other than the mean when counting low numbers, if the nuclei are distributed at random in space. This was diminished by taking many observations and on occasions when the curve was a straight line, representing slightly more than 80 % of the total, the error was probably less than ± 2°, ± 1°, ± 1° C for 0.1/l, 1/l and 10/l respectively. On the remaining occasions, including 4 out fo the 12 measurements for which more than 0.1/l was estimated at −10° C, the error may have been twice as great. Three 0.1/l estimates

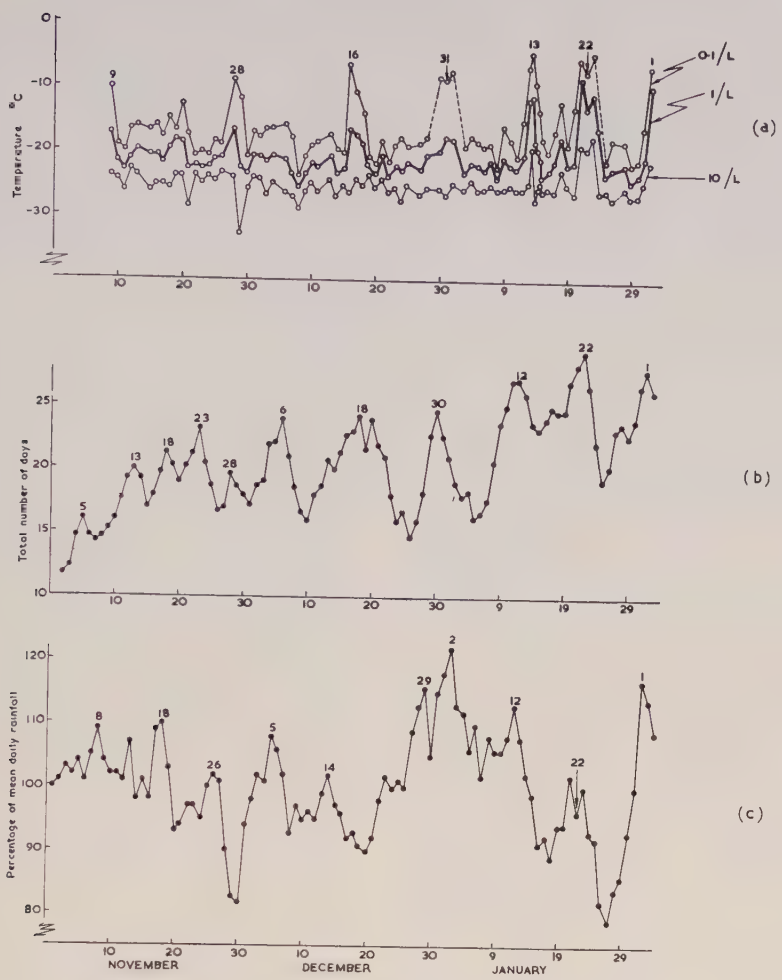


Fig. 3. (a) Temperatures of given freezing nucleus concentrations, Carnarvon 1955—56. (b) The frequency of occurrence of cirrus clouds over Western Australia, 6 stations for about 10 years. (c) Percentage of mean daily "world rainfall", after BOWEN.

were omitted from the graph because of high scatter, thought to be due to a patchiness in the distribution of the nuclei in the atmosphere. These were 27 November, 14 January (3rd reading) and 24 January.

(3) The possibility of spurious ice crystals from frost or nucleation in the detecting solution. Because standard solutions were used at fixed temperatures and a rigid procedure of glycerining the walls of the cold chamber was followed, it is not thought that errors from this source were serious.

Temperature was measured with an alcohol-in-glass thermometer accurate to ± 0.5° C over the range used. With its long time constant, and the possibility of incomplete equilibrium in the liquid, errors in measurement of the minimum cloud temperature may have been ± 1° C. As pointed out earlier, the mean temperature of the cloud was higher.

4. Results

Fig. 3a shows a graph of the temperature for three concentrations of ice crystals from 9 November 1955 to 1 February 1956 inclusive. Days missed, through lack of dry ice, were 14 November, 5, 12, 26, 29 December, 2 and 27 January. Two measurements were made

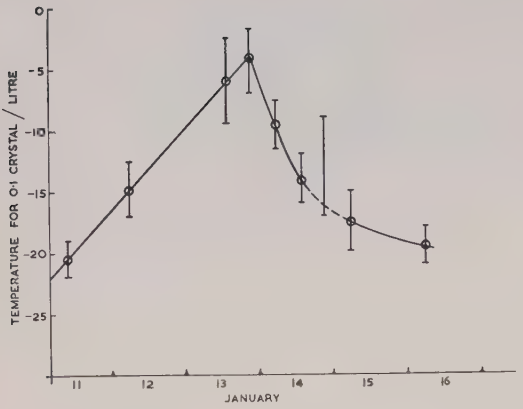


Fig. 4. Temperatures for 0.1 crystal/litre and estimated errors 11—16 January.

on 13 January and three on 14 January, but otherwise only one daily. It can be seen that there were small fluctuations from day to day interspersed with a few major changes, labelled in the diagram. The changes, which were most conspicuous at the 0.1/l level, were always seen in the 1/l count but only in January at concentrations of 10/l.

The major fluctuations were considerably greater than the estimated probable errors and represented systematic changes over several days. This is well illustrated by the measurements of 11—16 January when extra experiments were performed which are shown in Fig. 4. Here the limits of error shown were estimated only from the scatter of the points, thus ignoring systematic errors. The concentrations of nuclei active at −10° C present on 13 and 21—23 January and 1 February were as high as those commonly supposed to produce continuous rainfall, while on most other days this concentration was attained only below −20° C. Since their activity is comparable with that of silver iodide their influence on weather situations favourable to rain may have been considerable.

The average temperatures for concentrations of 0.1, 1 and 10/l were −16° C, −20.5° C and −25° C respectively. The temperature for 1/l was thus about 8° C higher than those found in Sydney and Arizona in January 1954 and 1955 by Smith and Kassander. This may be partially due to their measurements being made from aircraft and partially to the different equipment. The equipment described here, when operated in Sydney on the ground in 26 observations during September 1955, yielded a mean temperature of −18.7° C for 1 crystal/l.

5. Source of the nuclei

It has been reported by several observers that combustion residues, soil, or even microorganisms may act as freezing nuclei (see, for example, Soulage 1955) so that it is important to determine whether the observed fluctuations in ice crystal counts were due to purely local sources of nuclei. At the site chosen, air passed over varying amounts of land according to the wind direction, with a minimum of less than 1 km (consisting of a tidal swamp) for directions between SW—the prevailing wind—and NW.

Graphs of mean wind speed and direction during the period of observation were prepared and compared with the corresponding measurements of temperature for 0.1 crystal/l.

There was clearly no consistent relationship between the graphs and the inference is that the nuclei were not of local origin. The lack of correlation with wind speed does not favour the possibility of an oceanic source.

Combustion nuclei could have been only rarely present in the air sampled, the country being virtually uninhabited, except for the township to the east of the site. Notice that the mean temperature for 0.1/l in a smoky industrial suburb of Sydney in September was only 1.8° C higher than for Carnarvon in the period November to February. Evidently only a minute proportion of combustion nuclei has the property of being freezing nuclei, and their effect may be neglected at most sites.

No correlation with any other meteorological phenomenon was found during this period, with the exception of the occurrence of cirrus cloud. This type of cloud appeared, on the average, only on one day in three, yet it occurred on 11 out of the 15 days on which the temperature for 0.1 crystal/l was unusually high. Now, cirrus cloud will only form if a region simultaneously contains sufficient water vapour and nuclei capable of growing ice crystals at its temperature. If the relation observed is real and not fortuitous, it means that nuclei at the ground appear at

about the same time as at cirrus levels—probably higher than 6 km in this latitude. If the terrestrial source of nuclei is so remote that considerable vertical mixing has had time to occur, it is strange that large fluctuations should be observed on some occasions. An alternative explanation of the frequent simultaneity is that the nuclei are transported rapidly to low levels in ice crystals falling from the cirrus. The question of how they reached the cirrus in the first place then arises.

BOWEN (1953) has suggested that active freezing nuclei may be formed by dust accompanying (or resulting from) meteor showers and has summarized measurements (BOWEN 1956) which appear to support this contention. Cirrus may be regarded as an indication of the presence of active nuclei, for the more active the nuclei present, the more frequently will the conditions for formation of the cloud be satisfied.

Thus, according to BOWEN's theory, the frequency of occurrence of cirriform cloud should depend on date, and should be related to measured freezing nucleus concentrations. Of course, in any one year, the presence or absence of water vapour will also determine whether or not the cirrus forms and may influence the time of arrival of the nuclei at the ground. Consequently a similarity, rather than a complete correspondence between cirrus and freezing nucleus concentrations would be expected.

Records of the frequency of cirrus cloud occurrence over Western Australia have been compiled from the available records (about ten years') of six widely separated stations. These are shown in Fig. 3b as three-day means in order to compare them with BOWEN's (1956) published curve of rainfall, based on about 50 years' records of 300 places distributed round the world. This is shown in Fig. 3c.

Quantitative comparison of Figs. 3a, 3b, 3c through correlation coefficients may be rendered inexact by the fact that departures from the mean are not randomly distributed and each series has some serial correlation. MCDONALD (1957) has summarized arguments from the literature on statistics which show that errors arising from such causes are more likely to be of academic than practical interest when dealing with geophysical data. If the correlation coefficient, assuming normally distributed variables and no serial correlation were significant at the 99 % level, it would be immaterial that the rigorously derived significance figure should be 98 %. The point of making a quantitative comparison at all is to test whether one's opinion is due to bias, and as such is a useful procedure. In the author's opinion, Figs. 3a, 3b and 3a, 3c show a degree of association which would not be expected in unrelated series. The opinion is confirmed by correlation coefficients of 0.38 and 0.22 respectively, significant at about the 99.9 % and 95 % levels. Moreover, the Southern Hemisphere countries' contribution only to Fig. 3c (i.e. Argentine, Australia, New Zealand and South Africa) produces a curve which has the much higher correlation coefficient of 0.42 with Fig. 3a, significant at the 99.99 % level.

Thus, even allowing errors in these estimates of significance, there is undoubtedly far more agreement between the curves than is likely to arise from chance.

The implication is that the nuclei, the cirrus and the rainfall have some common control, and this is consistent with BOWEN's theory of a meteoric influence on rainfall.

A terrestrial source of nuclei is not ruled out. If the quoted results are typical the source must be distant, in order to explain the observed relation between phenomena at high and low levels and the lack of correlation with weather. The dependence of variations in the supply of nuclei on calendar date implied by the relation between curves including 1, 10 and 50 years' data is difficult to explain in terms of a terrestrial source. Note, however, that as there is not a one-to-one correspondence, the pattern is probably only partially repeated from year to year.

It will require many more such series of measurements at different sites before the source can be fixed with certainty. Desirable information which could be obtained by the method described is: the correlation between simultaneous measurements at sites separated by varying distances, the importance of local sources of nuclei in other parts of the world, and the vertical distribution of nuclei. A knowledge of the nature and size distribution of effective nuclei is clearly essential but will have to be obtained by other techniques.

6. Conclusion

The equipment described represents an advance on former techniques in giving an objective measurement of ice crystal concentration in a supercooled cloud. This should be as valuable in the search for useful cloud-seeding substances as in the measurements of natural nuclei. In the form used in these experiments it is extremely simple, reliable and easy to use, but for many applications larger working volumes, expansion techniques or other means of refrigeration have proved desirable.

The series of readings taken with it has shown that there occur large changes in the ice crystal concentration, often of short duration. The fluctuations are large enough to be a factor worth considering in weather modification programmes, and must have some bearing on whether or not natural clouds will rain.

The need for a world-wide series of observations of this sort is obvious, for conditions at this one site may not be typical. The problem of the source of the nuclei requires a similar approach if it is to be resolved.

7. Acknowledgements

I should like to thank the Director of Meteorological Services for providing the information on occurrence of high cloud, and the meteorological staff at Carnarvon airport for local data.

REFERENCES

Bowen, E. G., 1953: The influence of meteoritic dust on rainfall. *Aust. J. Phys.*, **6**, pp. 490—497.

Bowen, E. G., 1956: The relation between rainfall and meteor showers. *J. Meteor.*, **13**, pp. 142—151.

Cwilong, B. M., 1947: Sublimation in a Wilson Chamber. *Proc. Roy. Soc.* A, **190**, p. 137.

McDonald, James E., 1957: A critical evaluation of correlation methods in climatology and hydrology. University of Arizona, Institute of Atmospheric Physics, Sci. Rep. No. 4.

Smith, E. J., Kassander, A. R., and Twomey, S., 1956: Measurements of natural freezing nuclei at high altitudes. *Nature*, **177**, pp. 82—83.

Soulage, G., 1955: Sur la constitution et l'activité glaçogène des noyaux de congélation atmospheriques. *Comptes Rendus Ac. Sc.*, **240**, pp. 2168—2170.

Steady Convective Motion in a Horizontal Layer of Fluid Heated Uniformly from Above and Cooled Non-uniformly from Below

By HENRY STOMMEL and GEORGE VERONIS, Woods Hole Oceanographic Institution[1]

(Manuscript received March 28, 1957)

Abstract

A gravitationally stable fluid layer is heated non-uniformly (sinusoidally) at the bottom (or top) and a thermal convection regime is established. Examples are worked out for cases (i) without rotation (ii) with uniform rotation, and (iii) with non-uniform rotation (beta-plane).

1. Introduction

The study described in this paper is an outgrowth of a very interesting Russian paper by LINEYKIN (1955) in which he attempted to determine whether geostrophic currents, produced in a stable sea by wind, extend only to a certain limiting depth. There is, of course, ample observational evidence to indicate that currents decrease with depth because of the baroclinity of the ocean. The essential features which Lineykin incorporated into his model to limit the depth of currents were stratification of the ocean (in equilibrium, a linear increase of density with depth) and the Coriolis parameter (uniform rotation). The scale length of the model was taken to be 100 km so that his results could not be applied directly to large-scale oceanic motions. Evidently Lineykin had the Caspian Sea in mind, because attempts to extend his results to systems of horizontal oceanic scale lead to essentially barotropic (vertically uniform) velocities for depths of 5 km. It appears therefore that some additional physical factor—such as the variation of the Coriolis parameter—must be added in order to limit the depth of the currents much more drastically than the mechanisms envisaged by Lineykin. We have incorporated the variation of the Coriolis parameter into a simple model (§ 5) and have shown that it does indeed limit the currents to a depth which is much more realistic.

However, the basic objection to Lineykin's model and to the one discussed in this short paper is that the mean density structure of the ocean cannot be approximated by a constant vertical potential temperature gradient. Indeed, the problem appears to be essentially non-linear with advective processes playing a fundamental role in the formation of the main thermocline. For this reason we hesitate to offer the following results as a real explanation of the observed phenomena in the ocean.

Recently we attended a series of lectures on tropical meteorology by Drs. Herbert Riehl, Eric Palmén, and Joanne S. Malkus, and we realized that in some respects the atmosphere in the vicinity of the trade-wind regions is a more suitable phenomenon to which to apply Lineykin's analysis. In particular, it is much more nearly uniformly stable, and therefore, Lineykin's linearization procedure is more justifiable. It occurred to us that Lineykin's model might be formulated in such a way as to give a crude theory of the Trades. We are not qualified to judge whether the model is truly applicable to the Trades, but we have worked out the analytical details involved and present them here for the convenience of other workers. Also, of course, the results are in themselves of some interest.

2. Basis of the Model

Consider a fluid in a horizontal layer heated uniformly at the top and cooled below. Because of the thermal expansion of the fluid

[1] Contribution No. 899 from the Woods Hole Oceanographic Institution. This work was done under the auspices of the Office of Naval Research.

the density decreases upward, and the fluid is gravitationally stable. A uniform gradient of temperature $\partial\vartheta/\partial z$ (where z is directed positively upward and ϑ indicates temperature) is established; the heat flux $\varkappa\dfrac{\partial\vartheta}{\partial z}$ is determined by thermometric conductivity ($\varkappa$) alone and there is no fluid motion.

If the fluid is cooled non-uniformly below by a temperature perturbation $\vartheta_0 \cos ly$ impressed at the boundary $z=0$, then we may anticipate horizontal differences of density in the fluid. Buoyancy forces produce convective motions which must be maintained against viscous and conductive dissipation; hence the amplitude and vertical extent of the convection must be determined. The present communication endeavors to describe these convective motions in the case where the motions are very slow and the disturbances of the temperature field are very slight (as compared to the basic uniform temperature field); in other words, this is a linearized perturbation theory.

The perturbation equations which we shall use are:

$$-f\varrho_0 v = -\frac{\partial p}{\partial x} + A\varrho_0\frac{\partial^2 u}{\partial z^2} \tag{1}$$

$$f\varrho_0 u = -\frac{\partial p}{\partial y} + A\varrho_0\frac{\partial^2 v}{\partial z^2} \tag{2}$$

$$g\varrho = -\frac{\partial p}{\partial z} \tag{3}$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{4}$$

$$w\overline{\frac{\partial\vartheta}{\partial z}} = \varkappa\frac{\partial^2\vartheta}{\partial z^2} \tag{5}$$

$$\varrho = \varrho_0(1-\alpha\vartheta) \tag{6}$$

Equations (1) and (2) are the linearized steady state equations of motion for a rotating system (f is the Coriolis parameter). Since the vertical scale of the perturbation fields in the examples discussed is much less than the horizontal scale, there appears to be no reason to carry the full Laplacian in the frictional term and we have kept only the vertical viscous shearing stress term. Equation (3) is the hydrostatic equation. Equation (4) is the Boussinesq approximation to the conservation of mass equation. Equation (5) is the steady state perturbation heat flow equation linearized by taking $w\overline{\dfrac{\partial\vartheta}{\partial z}}$ (where the bar indicates the mean unperturbed thermal gradient) as the only important heat advection term. We suppose this to be balanced by the divergence of the vertical conductive perturbed heat transport. Equation (6) is a simple linear equation of state.

We now proceed to discuss solutions of these equations for those cases (§ 3) no rotation, (§ 4) uniform rotation, and (§ 5) non-uniform rotation.

3. No rotation

Consider solutions independent of the x-direction, and in which $f=0$. Eliminate all dependent variables except ϑ, thus obtaining the equation

$$\frac{\partial^6\vartheta}{\partial z^6} = -\frac{g\varrho_0\alpha b}{A\varkappa}\frac{\partial^2\vartheta}{\partial y^2}, \tag{7}$$

where we have introduced b to represent the basic temperature gradient $\overline{\partial\vartheta/\partial z}$. Consider a simple bottom temperature perturbation of the form $\vartheta=\vartheta_0\cos ly$, and a solution of the form $\vartheta=\Theta(z)\cos ly$, in which case equation (7) reduces to an ordinary differential equation

$$\frac{d^6\vartheta}{dz^6} = F^6\vartheta \tag{8}$$

where

$$F=\sqrt{\frac{g\varrho_0 abl^2}{A\varkappa}}$$

The quantity F^{-1} is the characteristic vertical scale of the convection field. We can call it the "frictional depth".

The solution is of the form

$$\vartheta=\vartheta_0\cos ly\sum c_i e^{p_i Fz} \tag{9}$$

where the p_i are the sixth roots of unity. If we consider only the three terms which become negligibly small for large z, we need two boundary conditions at $z=0$ in addition to the condition $\Theta(0)=\vartheta_0$. Physically it is clear that at the bottom $w=v=0$ are these conditions; in terms of Θ these two additional conditions are

$$\frac{d^2\Theta}{dz^2}=\frac{d^3\Theta}{dz^3}=0,$$

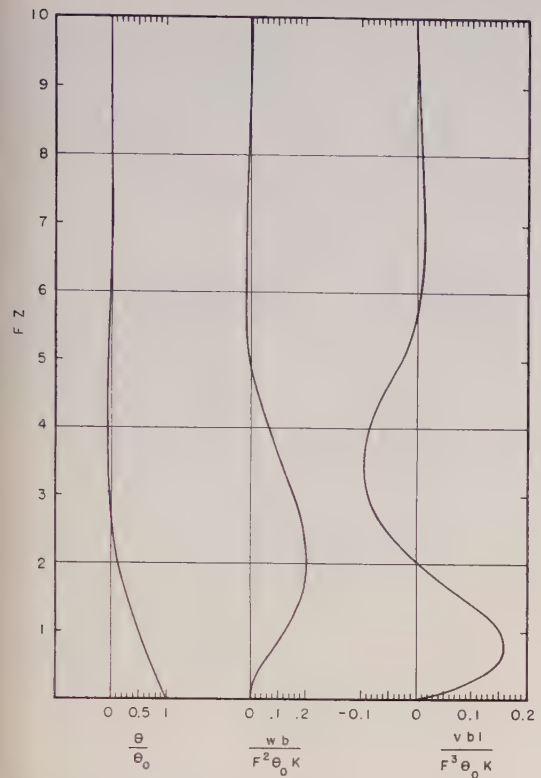


Fig. 1. Amplitude of perturbation temperature, and of the vertical and horizontal components of velocity as a function of altitude for the non-rotating case. Both coordinates are dimensionless.

from equations (5) and (4) respectively. The solutions of the problem are

$$\vartheta = \vartheta_0 \cos ly \left\{ \frac{1}{2} e^{-Fz} + \frac{1}{2} e^{-\frac{Fz}{2}} \left(\cos \frac{\sqrt{3}}{2} Fz + \right. \right.$$
$$\left. \left. + \frac{\sqrt{3}}{3} \sin \frac{\sqrt{3}}{2} Fz \right) \right\} \qquad (10)$$

$$w = \frac{F^2 \vartheta_0 \varkappa}{2b} \cos ly \left\{ e^{-\frac{Fz}{2}} + e^{-\frac{Fz}{2}} \left(-\cos \frac{\sqrt{3}}{2} Fz + \right. \right.$$
$$\left. \left. + \frac{\sqrt{3}}{3} \sin \frac{\sqrt{3}}{2} Fz \right) \right\} \qquad (11)$$

$$v = \frac{F^3 \vartheta_0 \varkappa}{2bl} \sin ly \left\{ -e^{-Fz} + e^{-\frac{Fz}{2}} \left(\cos \frac{\sqrt{3}}{2} Fz + \right. \right.$$
$$\left. \left. + \frac{\sqrt{3}}{3} \sin \frac{\sqrt{3}}{2} Fz \right) \right\} \qquad (12)$$

These relations are plotted in Fig. 1. The convective motion is in the form of long strip cells, or rolls, with axes in the x-direction. Fluid rises over warm portions of the bottom, and sinks over colder portions. The amplitude of the motion is proportional to the temperature perturbation at the bottom. The vertical extent of the motion is directly proportional to viscosity, conductivity, and horizontal scale. It varies inversely as the gravitational stability of the mean density field.

4. Uniform Rotation

In this case f is a constant and there is an x-component of velocity, u, but all variables are independent of the x-coordinate. Elimination of variables leads to an equation analogous to equation (8).

$$-\frac{1}{E^4} \frac{d^7\vartheta}{dz^7} + \frac{d^3\vartheta}{dz^3} - L^2 \frac{d\vartheta}{dz} = 0 \qquad (13)$$

where

$$E^4 = -\frac{f^2}{A^2}, \; L^2 = \frac{g\varrho_0 abl^2}{f^2} \frac{A}{\varkappa}$$

The quantity $2\pi/E$ is the well-known "Ekman depth" and the quantity L^{-1} can be called the "Lineykin depth".

In the case where the only terms admitted in the solution are those which vanish at infinity and $L \gg E$ the solutions are

$$\frac{\vartheta}{\vartheta_0 \cos ly} = 2e^{-\frac{E}{\sqrt{2}}z} \left[D_1 \cos z \frac{E}{\sqrt{2}} \right.$$
$$\left. + D_2 \sin z \frac{E}{\sqrt{2}} \right] + Ce^{-Lz} \qquad (14)$$

$$\frac{u}{\vartheta_0 \sin ly} = \frac{\sqrt{2} g\alpha l}{Ef} \left[(D_1 - D_2) \sin \frac{E}{\sqrt{2}} z - \right.$$
$$\left. - (D_1 + D_2) \cos \frac{E}{\sqrt{2}} z \right] e^{-\frac{E}{\sqrt{2}}z} -$$
$$- \frac{g\alpha l}{Lf} Ce^{-Lz} - \frac{\sqrt{2} A\varkappa E^5}{flb} \left[(D_1 - D_2)(\cos \frac{E}{\sqrt{2}} z + \right.$$
$$\left. + (D_1 + D_2) \sin \frac{E}{\sqrt{2}} z \right] e^{-\frac{E}{\sqrt{2}}z} +$$
$$+ \frac{A\varkappa}{flb} CL^5 e^{-Lz} + \text{const.} \qquad (15)$$

$$\frac{v}{\vartheta_0 \sin ly} = \frac{\sqrt{2}E^3\varkappa}{lb}\left\{(D_2 - D_1)\sin\frac{E}{\sqrt{2}}z - \right.$$

$$\left. - (D_1 + D_2)\cos\frac{E}{\sqrt{2}}z\right\}e^{-\frac{E}{\sqrt{2}}z} +$$

$$+ \frac{\varkappa L^3}{lb}Ce^{-Lz} \qquad (16)$$

$$\frac{w}{\vartheta_0 \cos ly} = \frac{2E^2\varkappa}{b}\left\{D_1 \sin\frac{E}{\sqrt{2}}z - \right.$$

$$\left. - D_2 \cos\frac{E}{\sqrt{2}}z\right\}e^{-\frac{E}{\sqrt{2}}z} + \frac{\varkappa L^2}{b}Ce^{-Lz} \qquad (17)$$

where

$$D_1 = \frac{L^2(2L^2 - E^2)}{4L^4 - 2L^2E^2 + 2\sqrt{2}LE^3 + 2E^4}$$

$$D^2 = \frac{L^2(E^2 + \sqrt{2}LE)}{4L^4 - 2L^2E^2 + 2\sqrt{2}LE^3 + 2E^4}$$

$$C = \frac{E^3}{E^3 - EL^2 + \sqrt{2}L^3}$$

These relations are plotted in Fig. 2. The Ekman layer is noticeable only in the lowest portion of the figure, the remainder of the profile being controlled essentially by the Lineykin depth. Qualitatively gravitational stability and horizontal scale have the same kind of influence on the vertical extent of the convective motion as in the non-rotating case, but their influence is stronger. On the other hand, the viscosity and conductivity enter as a ratio: if, in large-scale problems one were to suppose that they are to be interpreted as eddy-coefficients of comparable magnitude, the Lineykin depth must be largely *independent* of their amplitude. The ordinates are plotted non-dimensionally for Lz since L^{-1} is the controlling depth, but if E is changed the shape of the lower portions of the curves is modified. Thus the curves in Fig. 2 are not universal curves in the same sense as those in Fig. 1. The origin for the u-curve in Fig. 2 is arbitrary as indicated by the constant in equation (15). The constant can be determined by specifying the pressure at the level $z=0$. If one imposes $u=0$ at $z=0$, u approaches a constant value with height. If u is supposed to vanish at infinity it is finite at $z=0$. In either case, however, $\partial u/\partial z$ vanishes at $z=0$.

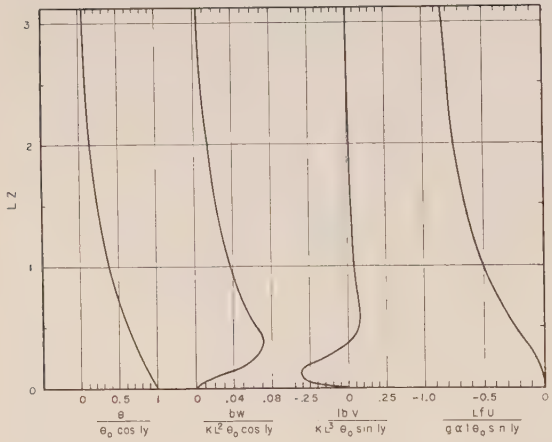


Fig. 2. Amplitude of the perturbation temperature and components of velocity as functions of altitude for the case of uniform rotation. The ordinate is the dimensionless quantity Lz expressed in terms of the Lineykin-depth. However, since there is an additional characteristic depth E^{-1}, it is not possible to present universal curves, particularly for the lowest values of Lz. The abscissae are dimensionless.

5. Non-uniform Rotation

Let us now consider the case where the Coriolis parameter varies with latitude, i.e., $\beta = \partial f/\partial y \neq 0$. For reasons that will become evident, the geometry of the problem is also changed; the perturbed boundary condition is taken as a function of x instead of y; at $z=0$, $\vartheta = \vartheta_0 \cos lx$; and also, in order to correspond to the ocean, the top instead of the bottom is taken at $z=0$. We are thus concerned with the solution in the region $z \leq 0$. It can be shown that for a certain choice of parameters (in particular large horizontal scale of oceanic dimensions) the main features of the temperature field come from the following approximate equation for ϑ:

$$\frac{\partial^4\vartheta}{\partial z^4} = \frac{\beta g\alpha b}{f^2\varkappa}\frac{\partial\vartheta}{\partial x} \qquad (18)$$

and if we let $\vartheta = \Theta e^{ilz}$ we obtain

$$\frac{d^4\Theta}{dz^4} = iG^4\Theta$$

where $G = \sqrt[4]{\frac{\beta g\alpha bl}{f^2\varkappa}}$

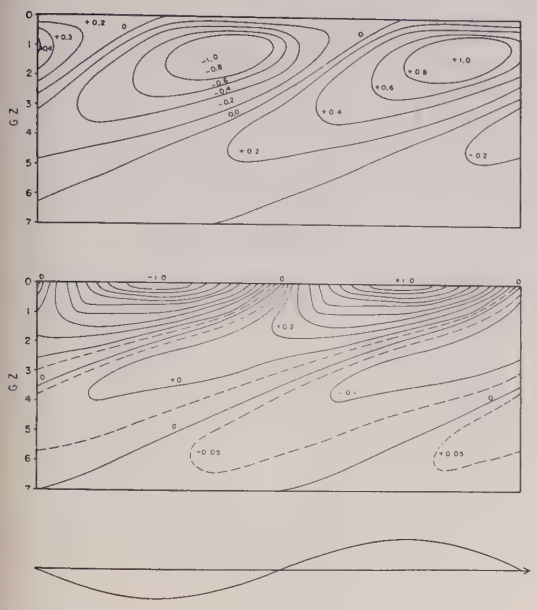


Fig. 3. (a) Contours of equal perturbation temperature (Ocean example with the non-dimensional depth Gz measured positive downward) in the case of non-uniform rotation. The horizontal coordinate is directed eastward toward the right of the figure. The perturbation temperature at the surface ($z=0$) is a sine function. The origin of the x-coordinate is in the center of the figure. The numbers on the contours represent the fraction of the amplitude of the surface perturbation.

$$2\sqrt{2}\,\frac{w f^2 r^2}{\beta g \alpha l T_0}$$

(b) Contours of the vertical velocity component in dimensionless form

The vertical scale in both (a) and (b) is exaggerated.

These equations can be obtained directly from equations (1)—(6) if the viscous terms in (1) and (2) are omitted. (It can be shown that for large-scale disturbances the effect of A is negligible below the Ekman layer.) The quantity defines a new characteristic depth which depends upon a different combination of parameters. There is a phase shift toward the west with depth, so that regions (or cells) whose perturbation temperature has the same sign tilt toward the west. Figures similar to Figs. 1 and 2 are not so easily interpretable. Therefore the full non-dimensional field of ϑ and w in the x—z plane is shown in Fig. 3, according to the solutions of equation (18):

$$\vartheta = \frac{\vartheta_0}{2}\Big\{\cos lx\Big[e^{-Gz\cos\frac{\pi}{8}}\cos\Big(Gz\sin\frac{\pi}{8}\Big)+$$

$$+e^{-Gz\cos\frac{3\pi}{8}}\cos\Big(Gz\sin\frac{3\pi}{8}\Big)\Big]+$$

$$+\sin lx\Big[e^{-Gz\cos\frac{\pi}{8}}\sin\Big(Gz\sin\frac{\pi}{8}\Big)-$$

$$-e^{-Gz\cos\frac{3\pi}{8}}\sin\Big(Gz\sin\frac{3\pi}{8}\Big)\Big]\Big\}. \quad (19)$$

$$w=\frac{\sqrt{2}}{4}\frac{G^2\varkappa\vartheta_0}{b}\Big\{\cos lx\Big[e^{-Gz\cos\frac{\pi}{8}}$$

$$\Big(\cos\Big(Gz\sin\frac{\pi}{8}\Big)+\sin\Big(Gz\sin\frac{\pi}{8}\Big)\Big)-e^{-Gz\cos\frac{\pi}{8}}$$

$$\Big(\cos\Big(Gz\sin\frac{3\pi}{8}\Big)-\sin\Big(Gz\sin\frac{3\pi}{8}\Big)\Big)\Big]$$

$$+\sin lx\Big[-e^{-Gz\cos\frac{\pi}{8}}\Big(\cos\Big(Gz\sin\frac{\pi}{8}\Big)$$

$$-\sin\Big(Gz\sin\frac{\pi}{8}\Big)\Big)+e^{-Gz\cos\frac{3\pi}{8}}$$

$$\Big(\cos\Big(Gz\sin\frac{3\pi}{8}\Big)+\sin\Big(Gz\sin\frac{3\pi}{8}\Big)\Big]\Big)\Big\} \quad (20)$$

Table 1.

Parameters	Experimental Water Model	Atmosphere	Ocean	Units
A	10^{-2}	$5\cdot10^{5}$	5	cm²/sec
K	10^{-3}	$5\cdot10^{5}$	5	»
$\frac{1}{\varrho}\frac{\partial\varrho}{\partial z}$	$2.5\cdot10^{-4}$	$2.5\cdot10^{-7}$	$5\cdot10^{-9}$	cm^{-1}
l	10^{-1}	10^{-8}	10^{-8}	cm^{-1}
f	1	10^{-4}	10^{-4}	sec^{-1}
α	$2.5\cdot10^{-4}$	$2.5\cdot10^{-3}$	$2.5\cdot10^{-4}$	°C^{-1}
β		$2\cdot10^{-13}$	$2\cdot10^{-13}$	$cm^{-1}\,sec^{-1}$
F^{-1}	0.4	$1.5\cdot10^{5}$	$9\cdot10^{3}$	cm
E^{-1}	0.1	$0.7\cdot10^{5}$	$2.2\cdot10^{2}$	cm
L^{-1}	7.0	$6.3\cdot10^{5}$	$45.0\cdot10^{5}$	cm
G^{-1}		$3.2\cdot10^{5}$	$0.46\cdot10^{5}$	cm

$$F^{-1} = \left[\frac{A\varkappa}{g \frac{1}{\varrho} \frac{\partial \varrho}{\partial z} l^2} \right]^{1/6}$$

$$E^{-1} = \left[\frac{A}{f} \right]^{1/2}$$

$$L^{-1} = \left[\frac{f^2}{g \frac{1}{\varrho} \frac{\partial \varrho}{\partial z} l^2} \quad \frac{\varkappa}{A} \right]^{1/2}$$

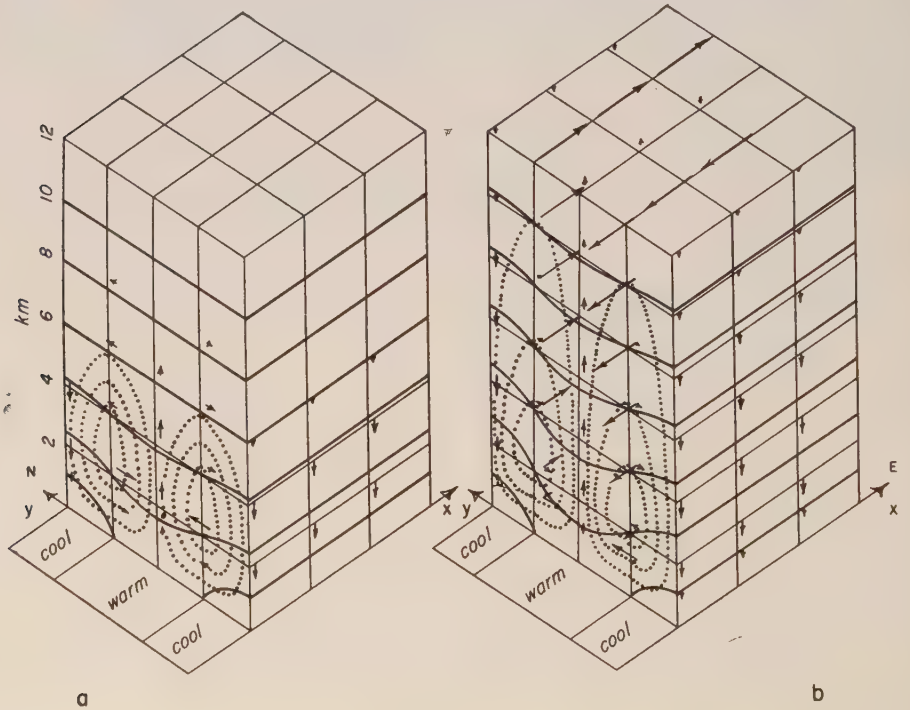
$$G^{-1} = \left[\frac{f^2 \varkappa}{g \frac{1}{\varrho} \frac{\partial \varrho}{\partial z} l\beta} \right]^{1/4}$$

6. Discussion of Results

In order to exhibit the contrast in temperature and velocity structure in the different cases, parameters appropriate to three different physical situations are inserted into the expressions for the various characteristic lengths and the results tabulated in Table I.

The water experimental model is envisaged as being about 10 cm deep, with a basic vertical temperature gradient of 1°/cm. The bottom temperature is perturbed with a wave length of 60 cm, and an amplitude of not much more than 0.1° C (otherwise the perturbation approximation breaks down). The viscosity and conductivity coefficients are molecular. Under these conditions a very thin ($F^{-1} = 0.4$ cm) convective cellular motion arises with rising over the warm areas of the bottom and sinking over the cool areas, and compensating horizontal flows. As can be seen from Fig. 1, the top of the cells is effectively at $Fz = 6$, or about 2.4 cm. If this whole system is uniformly rotated at about one revolution every twelve seconds, the height of the cellular pattern increases about seventeen times, a thin Ekman layer appears close to the bottom, and a large component of geostrophic flow transverse to the horizontal temperature gradients appears. As it is not easy to picture how a non-uniformly-rotating experimental model could be constructed this case is not discussed.

The parameters in the atmospheric case are chosen to coincide as nearly as possible to



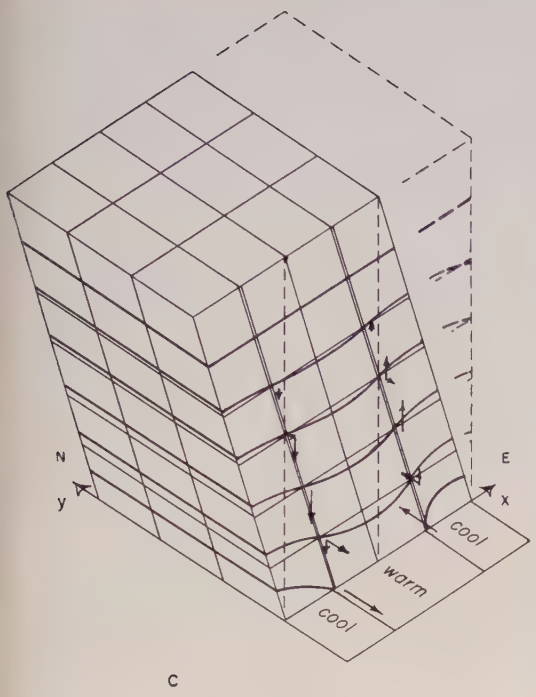


Fig. 4. Three cases using parameters for the atmosphere as given in Table I, with the vertical scale greatly exaggerated. The velocity components at selected points are shown by arrows. The heavy lines delineate iso-thermal surfaces. The dotted curves are the two-dimensional streamfunction in the y—z plane. Case (a) is non-rotating, there is no x-component of velocity as there is in case (b) of uniform rotation. In the case of non-uniform rotation (c) the applied boundary temperature is a function of x (instead of y as in (a) and (b)). The dashed lines represent the vertical, and the regions of perturbation temperature of same sign are enclosed by the tilted lines. Thus in (c) each warm and cold cell is tilted toward the west. On the other hand there is no x-component of velocity: all motion is confined to planes x-constant.

those used by Malkus and Riehl in a discussion of the Trades. The extremely large eddy coefficients are meant to include effects of cumulus convection as turbulent eddies. The mean stability corresponds to a vertical potential temperature lapse rate of 1°/100 m. The perturbation wave length is 6,000 km. The characteristic lengths are computed in Table I, and a three-dimensional sketch of the three atmospheric cases is shown in Fig. 4. Uniform rotation (b) produces a much higher cell than no rotation (a), and introduces zonal winds. The model for non-uniform rotation, in which the perturbed bottom temperature is a function of longitude (4c) instead of latitude, has a marked westward tilt with height. There is no horizontal component of cross-isobar flow because in the simplified case treated in Section 5 there is no friction. All flow patterns are geostrophic and in vertical meridional planes. The vertical velocity components are associated with the planetary divergence of the meridional winds.

The oceanic model in Table I is very unrealistic on account of the absence of meridional barriers. It is interesting to note that lack of the rotation gives a very shallow depth of penetration of the perturbation temperature (or broadly speaking, depth of the main thermocline) (90 meters); in the uniform rotation case the Lineykin depth of the current is 45 km; non-uniform rotation (c) leads to a thermocline depth of 470 meters.

REFERENCES

LINEYKIN, P. S., 1955: On the determination of the thickness of the baroclinic layer in the sea. *Doklady SSSR Akad. Nauk*, Tom 101, p. 461.

Proper and Improper Use of Statistics in Geophysics*

By BLAIR KINSMAN, Chesapeake Bay Institute, The Johns Hopkins University, Baltimore, Maryland

(Manuscript received November 23, 1956, revised March 22 1957)

Abstract

A discussion of some of the peculiarities of the geophysical sciences together with their implications for statistical methods, particularly for the correlation coefficient, is given. The discussion is illustrated by an analysis of a recently published paper. In conclusion some advice on the use of correlations is offered.

Introduction

The notion that the business of a scientist is the telling of "likely stories" is at least as old as Plato's "Timaeus". The materials with which a scientist works are "appearances", that is sense impressions, and in the vivid old Greek phrase, he tries to "save the appearances". This is a fairly straight forward job. An excellent example is offered by Ptolemy, who takes astronomical observations back to the reign of Nabonasser and shows that he can "save" them, that is fit them into a coherent pattern, by telling either a geocentric story or a heliocentric story. With two equally satisfactory stories Ptolemy weighs their likelihood and, on the basis of the physical information available to him, develops the geocentric story as the more likely. The basis for this choice is hardly so simple and straight forward. Today the general consensus of opinion, except possibly among navigators, is that the heliocentric story is the "true" one.

The advent of the word "true" together with the word "real" in scientific discussion has done much to cloud the nature of scientific activity. The request for a "true" story instead of for a "likely" story tacitly postulates the existence of a "real" world underlying and giving rise to appearances and asks for information about that "real" world. The scientist *qua* scientist cannot answer such a question since the material on which he operates consists entirely of appearances. St. Thomas Aquinas would probably have said that no one could answer since "Nothing was ever in the mind that was not first begot upon the senses". The attempts to bridge the gap between the postulated "real" world and the world of appearances which we perceive has a long and uniformly unsatisfactory history covering the spectrum from Descartes' assertion that God would not fool us to Berkeley's retreat into solipsism. If we restrict ourselves to "appearances", "hypotheses", and "likelihood", metaphysical speculation about "truth" and "reality" can be left to the metaphysician with a considerable gain in clarity.

The point of departure is the appearances. These range from the casual impressions of any sentient being through the systematic observation of essentially uncontrollable phenomena characteristic of geophysics to the

* Contribution No. 31 from the Chesapeake Bay Institute.

precise measurement of the results of highly controlled experiments characteristic of the laboratory sciences. The habit of attentive observation, coupled with an overwhelming urge to fit the observations to a pattern, embryonic in most of the human race, is developed in the scientist to a high degree. All appearances, however, are not suitable for scientific activity. Aristotle said that the subject matter of science is that which happens always or for the most part. The unique event is a subject for history. POINCARÉ (1905) puts it this way: "Carlyle has written something after this fashion. 'Nothing but facts are of importance. John Lackland passed by here. Here is something that is admirable. Here is a reality for which I would give all the theories in the world.' ..... The physicist would most likely have said: 'John Lackland passed by here. It is all the same to me, for he will not pass this way again.'"

Having, then, a set of observations of a recurring phenomenon the next step is to construct an intelligible hypothesis into which the observations can be made to fit. One fertile method is the use of analogy. Some other set of phenomena and their pattern being known, if we can see a similarity, we can transfer the properties of the known system to the unknown. Since analogies are seldom isomorphisms, the correspondences being only partial, the dangers of argument by analogy are obvious. For example, the complex numbers are analogous to the real numbers in the sense that operations of addition, subtraction, multiplication, and division can be defined for each. We might then argue by analogy that, since division by zero is not permitted for the reals, division by zero is not permitted for the complex numbers. We thus reach a correct result. If we argue by analogy that, since the reals are ordered, the complex numbers must also be ordered our conclusion is false. Fertile as the argument by analogy is as a source of ideas, it is almost worthless in support of an hypothesis unless it is shown that the analogous systems are similar in every essential feature and that no dissimilar features can affect the properties that we wish to establish. Another method is to search the data for regularities. If the sample is small (and this is often the case in geophysics) this is quite easy to do. However, since even samples drawn from a random number table will exhibit regularities, results from such a procedure are suspect and once a regularity is found most scientists feel impelled to provide some rationalization for it, often based on an *ad hoc* selection of arguments. (The indiscriminate use of the correlation coefficient is particularly offensive in this connection and we will have something to say about it later.) Perhaps the most acceptable method of forming hypotheses is by rational argument from established elementary principles. It is worth noting that scientists in general seem to feel uneasy about any hypothesis until it has been presented in this form no matter how it was first conceived. An argument in Kepler's "Epitome" is a case in point. Kepler's hypothesis that the planetary distances are governed by the proportions of the five regular or Platonic solids seems a little wayward today, but the urge to order the welter of appearances is easily understandable. Kepler apparently felt the need to deduce his hypothesis from the first principles of geometry and astronomy since he devotes considerable space to the effort. His logic is impeccable except for one short section. There the line of the argument has been blurred, whether consciously or unconsciously cannot be known, so that the ostensible deduction he was at such pains to make is, in fact, not established. The expenditure of so much effort in such a cause by a man with first rate discoveries to his credit shows the importance attached to this method.

But the telling of tales is only half the job and the easiest half at that. We still face the problem of deciding how likely the story is or, if confronted with two different but adequate stories, which is the more likely of the two. Statistics has been increasingly concerned with understanding the structure of such decisions and with finding a clear and objective method of making them. The general problem is far from solved but many valuable results have already been obtained. The judgments of the likelihood of an hypothesis have had so many different bases that even a simple enumeration would be too long. They include decisions made on entirely extraneous grounds, e.g. the selection of a flat earth over a round earth on arguments derived from the "second coming". They include decisions made on what I should call

aesthetic grounds, e.g. the selection of uniform circular motion as basic by the Greeks in contrast to uniform rectilinear motion by Newton. They include decisions in which maximum simplicity is equated with maximum likelihood. Occam's Razor is still a widely used scientific tool although the simplicity of nature is more an article of faith than a proven fact. The more we refine our studies of nature the more complex things become until it almost seems as though simplicity were an attribute of the infancy of a science. Fortunately, there are a few threads to guide us in the labyrinth. Occam's Razor has been mentioned. If you are willing to commit yourself to the proposition that the relations among phenomena are fundamentally simple, then you will always choose the simplest story that explains all the facts. However, it must always be borne in mind that one man's simplicity may be another man's utter confusion. A heliocentric hypothesis simplifies the astronomer's calculations but it makes those of the navigator intolerably complex. If one is willing to forego questions of "truth" and "reality" one can escape the dilemma by accepting both the geocentric and the heliocentric hypotheses (so long as they are not in logical conflict) and use whichever one is simpler *for the immediate purpose*. A most important criterion is that of compatibility with already existing structures. If the hypothesis under consideration would require extensive revision of major parts of a successful existing theory with all the labor that entails, clearly one would hesitate to accept it unless a general improvement throughout the whole theory could be anticipated. Another equally important basis for decision is the continued agreement of observations with the hypothesis. Here statistics enters since no set of experimental measures, if suffciently refined, ever agrees exactly with an hypothesis or with other sets. The tincture of statistics that most of us retain from our formal training seldom goes beyond the memory of where to find a few computing formulae. This isn't enough. While a sprinkling of Pearson's x^2's, Student's t-tests, and Fisher F-tests do lend an air of objectivity to any paper, it must be remembered that statistical tests of significance derive from mathematical models which in turn are based on different views of the nature of phenomena. The calculation of parameters is routine and their use often obscures a lack of precise thought about the fundamentals of a problem.

In general, the job of the scientist is to invent a story which accounts for a set of observations and then to decide how likely the story is.

Geophysics

The geophysical sciences share the same theoretical base with the laboratory sciences. The Navier-Stokes equations hold equally well for a beaker of water and for the oceans. With this common base it is not surprising that the methods for treating geophysical data are often selected by analogy with those used on laboratory data. Unfortunately, the materials on which these methods are used in geophysics are sufficiently different from those of the laboratory sciences to require justification of the method, which is seldom explicitly given. In the first place, the equations which give a full description of any situation are usually too complex to be handled in their complete form. It is almost always necessary to simplify them by considering some terms as negligible in order to get an approximate solution. The laboratory scientist by controlling the conditions of an experiment can insure that terms considered negligible are so in fact and, as a result, he can expect good agreement between hypothesis and experiment even with relatively small samples. Further, he can repeat his experiment at will.

The geophysicist who, in the main can only observe, cannot repeat his observations in the sense of a repeated experiment. This is a grave difficulty since most statistical tests of significance are fundamentally rooted in the idea of repetition. While such repetition is conceptually possible in geophysics it is seldom realized and, in general, when such tests are used, their use should be supported by argument. Further, his inability to insure that factors considered negligible are so should logically force the geophysicist to use the equations for a process in their complete form or first to show that the neglected factors can be neglected. Neither of these courses is usually taken. There is nothing wrong with the making of simplified models so long as they are not offered as "reality". When the

observations happen to agree with such models it is cause for gratification and suggests that the neglected terms were small. If the observations continue to agree with the model we can feel that we have a satisfactory story. In geophysics the agreement usually need not be very close before verification is claimed. In one case known to the author the originator of a widely accepted hypothesis will take a discrepancy of $\pm$ 20 % in the observations before he considers it worth any attention. This attitude stems from lack of control. In comparing data with a simplified model large dispersions are to be expected. This means that large samples are necessary if relations are to be established with any certainty. It is unfortunate when the investigator, having been forced to accept a simplified model, then feels impelled to insist that the natural phenomena are themselves simple.

Another difference arises from the answer that can be given to the question: "Do the data describe the phenomenon under consideration." In the laboratory sciences methods can usually be devised either to measure a property directly or to measure some closely linked property. With the method in hand, the experimenter can then accumulate enough data for a statistically reliable estimate of the property which interests him. The geophysicist here labors under two kinds of handicap. First is the matter of scale. Both the space and time scales are usually unwieldy. If one asks for the monthly average temperature of the Chesapeake Bay is it enough to dip a thermometer in once a day at some convenient place? One can hardly say without knowing a great deal about the structure of the Bay. If it were enough for, say, January 1949, would anything be known about January 1950? To answer such questions an inordinately expensive observation network would have to be established and maintained for many years. Salinity records taken daily at Solomon's Island, Maryland are a case in point. This set of data extending back to 1938 is the longest unbroken record of salinity taken anywhere on the Bay. Using monthly means and computing the power spectrum it was found that there was evidence of a yearly cycle, which was to be expected. However, the great bulk of the power in the signal occurred at periods greater than two years. It was calculated that to separate cycles having periods of three years and four years at the 5 % level, the record would have to extend over 285 years. Such a sample from a geophysical point of view is huge and it is rather sobering to see how little information it gives. Second, the geophysicist must frequently work with data which were taken for other purposes and which do not directly measure the properties that interest him. For example, an oceanographer interested in the factors influencing the size of fish populations might not have any measurements directly made for that purpose but instead measurements of salinity made at some point in the region and records of commercial fish catches. It isn't what he wants. It's what he's stuck with. If he persists, he would have to argue something like this. Fish population controls fish catch. Fish need plankton for food. Plankton need dissolved nutrients. Nutrients are brought to the surface layers by upwelling. Upwelling influences salinity. The salinity of the region can be determined from the salinities measured at a point which I know. Therefore, I will look over fish catches and salinities for possible correspondences and, if I can construct one, I will know the connection between fish population and environment. Laying aside the questions of whether the salinity measured at a point represents the salinity over a large area and of whether fish catch is an adequate measure of fish population, it seems unlikely that a definite clear cut relation between the ends of such a long and tenuous chain would emerge from a small sample.

Another hazard inherent in using existing data taken for other purposes arises from the temptation to fill gaps in it. If properties A and B are to be related and it is found that A was measured at some point for a number of years but that the measurement of B was neglected for a part of the time, then the urge to use measurements of B made somewhere else to fill the gap may be almost irresistible. This procedure enlarges the sample with an apparent increase in statistical reliability but it introduces tacitly the very difficult additional problem of showing that the measures introduced to fill the gap are the equivalent of what would have been secured had B been measured at the point. This is usually impossible. In using existing data for purposes

for which they were not taken great care must be exercised to see that wishful thinking does not govern the make-up of the sample.

Geophysics needs large samples. It has small ones. In contrast with the laboratory sciences these small samples are often imprecise, having been painfully secured in the field over many years, sometimes two or three generations. To get another sample for testing involves the same long process. Thus, if the entire initial sample is used in the formulation of an hypothesis we are forced to leave its verification or rejection to our grandsons. It is clear that progress of a science which must either proceed on untested hypotheses or wait for generations to test them will be either insecure or very slow.

Verification is to be had only from data not used in formulating an hypothesis. One possible method of securing data for testing a geophysical hypothesis formulated by search, without delay, is to split the data on hand into two groups, one to be used in formulating the hypothesis and the other reserved for testing. This may be done in a number of ways. In some fields dealing with time series every other time unit is grouped to form the two sets, or the data may simply be split in the middle. Separation by means of some randomizing device could be used so that the bias, conscious or unconscious, of the investigator would not invalidate such statistical tests of significance as might be appropriate. The real difficulty here is "keeping the game honest". If the hypothesis is formed before the test data are taken no question of influence arises. With both sets of data in existence at the beginning of an investigation there is always the question of the extent to which the investigator is influenced in his selection of hypotheses by the test data. A glimpse of it, however fleeting, could bias him toward hypotheses likely to fit both sets. The difficulty could be met if the separation were made before the investigator saw the data and he inspected only one set until he was ready to test his hypothesis. Any alterations in the hypothesis after testing would, of course, be highly suspect. The advantage of this device is that verification can be carried out at once and an estimate of the value of the hypothesis made. The disadvantage is that the already small sample size is further reduced, but it may be worth accepting this reduction in exchange for immediate evaluation. It is well to remember that the information contained by any finite sample is limited. Manipulating it in this way cannot increase the amount of information contained. It can only sacrifice information of one kind to gain information of another.

Correlation

A fallacy that is still far too prevalent in geophysical papers is the inference from a high correlation to cause and effect. Such inferences, when they are made, are usually concealed under a mass of verbiage and intensive critical reading is usually necessary to reveal them for what they are. The existence of a cause and effect relationship is never a necessary inference from a correlation.

The exhibition of high correlations between unlikely things provides considerable amusement deriving from the absurd inferences that suggest themselves. One of the author's favorites is the high correlation between increases in school teachers' pay and increases in the production of alcoholic beverages from which one may conclude that the way to promote sobriety is to cut teachers' salaries. COHEN (1944) gives an interesting discussion of causality and correlation beginning on page 132 of his "Preface to Logic" in which he cites a correlation of 0.87 between the membership of the International Machinists Union and the death rate of the State of Hyderabad.

Correlations may be used legitimately for exploratory work. If an investigator suspects for any reason that properties A to D may have a causal connection with X and wishes to fix on the most likely ones for immediate study, the computation of the correlations may be used as a guide. Suppose that A is highly correlated with X and that none of the others are. Then a thorough study of the relation between A and X is an obvious first step. However, just from the nature of correlations the investigator may be misled. The size of a correlation in itself proves nothing. It is merely a signpost pointing the way and it may very well be pointing the wrong way. The term "high" correlation has been used without saying what is meant. In some fields a correlation greater than 0.60 seems high

enough to cause excitement. How good is a correlation of 0.60? If we compute the coefficient of alienation we get 0.80 meaning that in comparison with a zero correlation, a correlation of 0.60 has reduced the standard error of estimate only 20 %. Eighty percent of the scatter still remains. The correlation must rise to 0.866 before the standard error of estimate is halved and even with a correlation of 0.99 the standard error of estimate is still 14 % of the standard deviation. As usually used in the literature, "high" seems to mean no more than "higher than we usually get with this kind of thing". It should be obvious that a correlation coefficient is not high for prediction purposes unless the major part of the variation is accounted for. As a rule of thumb one should not take a correlation seriously for prediction purposes unless it is 0.95 or higher, since it will then account for 70 % or more of the scatter *in the sample for which it was computed.*

The estimates of the true correlation in a universe computed from samples are strongly influenced by the size of the sample. If the true correlation is ϱ then the standard deviation of the sample correlations, r, is approximately $(1-\varrho^2)/\sqrt{n-1}$. Thus, if n is small quite large departures of r from ϱ are fairly common. If $\varrho = 0$, then an exact test based on the t-distribution is $t=(r\sqrt{n-2})/\sqrt{1-r^2}$ with $n-2$ degrees of freedom. In dealing with small samples the greatest caution should be exercised in accepting a high sample r as evidence of a high population ϱ.

If correlations are to be computed, the data should be used without preliminary modification. Failure of the data to measure the properties under consideration will act to lower any correlation that may actually exist but, however uncertain are the original data, no extra information can be forced into them by sitting at a desk and manipulating them. Occasionally one encounters a case in which the data have been first rounded off. The process of rounding off is sometimes useful as a means of expressing the degree of doubt about the original data, but calculating correlations with rounded figures may lead to biased and imprecise results. Less often one finds the data grouped and correlations computed between arbitrary numbers designating the groups. The worst of all practices, however, is the artifice of adjusting the data to improve the correlation. If the values are adjusted, the correlation can be improved to almost any extent but it then has little bearing on the phenomena the data purport to measure.

Another practice sometimes used with a pair of vaguely sinusoidal records such as frequently derive from meteorological phenomena is to slide them past each other until a fair correlation is obtained. Further lags are then investigated until another fair correlation is obtained and this is then used in connection with the first one to improve the agreement. Since this can always be done with truncated functions of this sort, the improvement is probably factitious.

If a correlation were computed and tested and allowed to stand or fall by itself the situation would not be too confused. However, one often finds that whole series of correlations have been computed. In general when a number of possible correlations are available it is not very fruitful to fix attention on one impressive one and ignore the others. This process of selection is not always deliberate. The chapter on ESP and PK in GARDINER (1952) illustrates how it may operate without any intention to deceive on the part of the investigator.

The correlation coefficient computed for a sample, then, is evidence for a relation obtaining within the sample for which it was computed. It is an estimate of an unknown conceptual true correlation existing in the universe from which the sample was drawn but it is an estimate that may easily be far wide of the mark. It is certainly no convincing evidence of any cause and effect relation. From the discussion, it should be clear that while correlations can be useful as supporting evidence, they can never play the role of primary arguments. In fact, the entire field of geophysics would be spared a great deal if every author would withhold from publication any paper devoted mainly to presenting arguments constructed to account for a high correlation.

Icebergs

The foregoing general discussion had its genesis over several years during which time critical reading of many papers in geophysics

and allied fields produced the uneasy feeling that quite a few of them contained less than met the eye. It was precipitated specifically by a paper by Louis A. Post (1956) called "The Role of the Gulf Stream in the Prediction of Iceberg Distribution in the North Atlantic".

After stating the fact that the number of icebergs south of 48° N per year varies widely, Post rejects the commonly held opinion that the causes are many and related in a complex fashion in favor of the notion of a single primary cause. He states his objective as follows: "It is the purpose of this paper to provide evidence that fluctuation in the Gulf Stream System is such a single cause and to derive a prediction technique for icebergs south of latitude 48° N from causative variates in this system."

We consider first the argument for a primary cause. When this is disposed of we will still have to look into the effectiveness of the prediction system whether the cause exists or not, since it is conceivable that one variable may be a good predictor of another even though they are not causally related.

Post's argument may be summarized as follows:

I. The Gulf Stream System, together with the Irminger, West Greenland, and Labrador Currents, may be taken as analogous to a closed circulatory system, information secured at one point being applicable to other points at other times.

II. The mean monthly temperature anomaly of the Florida Current at Key West[1] is inversely proportional to the "strength" of the current at Key West.

III. The "strength" of the Gulf Stream off the Grand Banks varies inversely as the temperature anomaly at Key West after a suitable delay.

IV. As a result of I through III, the number of icebergs south of 48° N varies directly with the anomaly.

V. After leaving the Grand Banks some of the water from the Gulf Stream is carried into the iceberg calving area in Baffin Bay by the Irminger and West Greenland Currents. Here a strong influx of warm water increases the number of icebergs calved so that a strong Gulf Stream current will increase the number of icebergs available for transport by the Labrador Current. It also serves to strengthen the Labrador Current so that after a suitable lag a larger number of bergs appears in the Grand Banks area. Since a strong Gulf Stream is associated with a low anomaly, the variation in icebergs with anomalies is inverse, allowing for a sufficient time lag to permit the current to reach Baffin Bay and return.

The analogy with a closed circulatory system (I) provides an appealingly simple picture. However, the absence of restraining walls in the Gulf Stream System makes it unclear that effects are transmitted through it in a manner analogous to their transmission through a closed system of tubes.

In the absence of velocity or mass transport measurements at Key West during the period under consideration, the argument in support of assertion II, that Key West temperature is inversely proportional to Key West current "strength", must be indirect. The data used as a test are the counts of icebergs found south of 48° N made by the International Iceberg Patrol[2] and the mean monthly temperatures at Key West (United States Department of Commerce, Special Publication No. 278, Revised (1951)). A correlation of 0.65 is found between the March anomalies and the iceberg count for 26 pairs of measures. This is evidence of a sort for a direct variation of the number of *icebergs* with the March anomaly. It is worth noting that the iceberg counts on which the correlation is based are highly subjective and have been made by many different people. In addition there is reason to suppose that the very low counts recorded for 1940 and 1941 may have been the result of lack of observers rather than lack of icebergs. If a connection does exist the error in the iceberg count would act to decrease the estimate of the correlation coefficient. If one accepts Post's contention that the number

[1] Hereafter, the word anomalies will denote mean monthly temperature anomalies of the Florida Current at Key West unless otherwise specified.

[2] The primary sources for iceberg counts are The Bulletins of the U.S. Treasury Department, Coast Guard, "International Ice Observation and Ice Patrol Service in the North Atlantic Ocean". Bulletin No. 21, the Season of 1931, tabulates counts from 1900—31. Bulletins Nos. 22—34 cover the years from 1932—48. The counts for years subsequent to 1948 were furnished to us by the U.S. Coast Guard, Washington, D.C. at our request. Whether these are the counts used by Post (1956) is not known.

of icebergs varies inversely with the "strength" of the Gulf Stream, it follows with the same force that the "strength" of the Gulf Stream varies inversely as the March anomaly. The step from the "strength" of the current off the Grand Banks to the "strength" of the current in the Straits of Florida is not very firm. Two arguments are offered. One, based on ISELIN (1938, 1940) holds that positive anomalies are associated with a stronger Florida Current which brings about a contraction of the North Atlantic Eddy. This in turn is accompanied by weaker peripheral currents in the region of the Grand Banks which permit the Labrador Current to carry its load of icebergs farther to the south thus increasing the number counted. The other, arguing by analogy with a closed system, holds that weak currents are transmitted directly through the system so that weak currents in the Gulf Stream imply weak currents in the Straits of Florida. The second argument is preferred. The assertion is thus seen to hinge on a correlation coffiecient, an assumed relation between Gulf Stream "strength" and numbers of icebergs, and an argument by analogy. The reasons for the selection of the March anomalies will be discussed later, but it should be noted here that three other months show higher correlations, July and December of the previous year, 0.72 and 0.70 respectively, and January, 0.76. Too much weight should not be given to the differences between these correlations as the sample size is not large enough to rule out substantial chance fluctuations. In the absence of data the assumed relation between Gulf Stream "strength" and numbers of icebergs is unsupported. The argument by analogy, while a useful guide, cannot serve as a secure foundation for a theory of causation. The contribution of the Antilles Current to the Gulf Stream has not been considered. A substantial part of the mass transport of the Gulf Stream has been contributed north of the Florida Current. Fluctuations in this additional flow might not agree with those of the Florida Current and the fluctuations in the Gulf Stream might be markedly different from those of the Florida Current. There is little direct evidence. Two studies by WERTHEIM (1953, 1954) provide some indirect evidence that the action of the Antilles Current may frequently be that of a damper on the Florida Current.

Assertion III, that Gulf Stream "strength" off the Banks varies inversely as Key West temperature anomaly, again goes back to the correlation. It is assumed that a weak Gulf Stream current off the Grand Banks permits the Labrador Current to carry its load of icebergs farther to the south while a strong Gulf Stream sweeps the icebergs northward out of the area. If this is granted, then the correlation of 0.65 between March anomalies and iceberg numbers implies an inverse variation of the Gulf Stream "strength" with temperature anomaly at Kay West.

The conclusion, IV, that the number of icebergs south of 48° N varies directly with the March temperature anomaly at Key West, is not a result of the previous steps. It is a direct inference from the correlation of 0.65. From the fact that the Gulf Stream system is the major source of warm water in the region of the Grand Banks it is reasonable to expect some connection between temperature fluctuation at Key West and iceberg numbers. That the variation is direct is suggested by the fact that the correlations for each of the twelve months is positive.

Table 1. Correlation of Iceberg Numbers and Mean Monthly Temperature Anomalies at Key West

Month	June	July	Aug.	Sept.	Oct.	Nov.
Correlation ..	0.31	0.72	0.08	0.07	0.17	0.10
Month	Dec.	Jan.	Feb.	Mar.	Apr.	May
Correlation ..	0.70	0.76	0.15	0.65	0.30	0.40

That the temperature anomaly at Key West is a primary cause is unproven. Many phenomena show fluctuations similar to the anomalies at Key West and could be made, by the type of selection used by Post, to show correlations with iceberg numbers as high or higher. In casting about for something so remote from iceberg numbers that there would be no temptation to fabricate an *ad hoc* argument for causation the author hit upon the notion of correlating the typeset commas per page in Post's paper as it was printed in Tellus. After a few minutes of exploring it was found that by pairing pages 111—102 with iceberg counts from 1942—1951 a correlation of – 0.81 resulted. This is considerably better than 0.65 but no one would argue that commas cause

icebergs. GARDINER (1952) gives an interesting popular discussion of the ease with which such striking correspondences can be manufactured in his chapter on the Great Pyramid.

The selection of the March anomalies needs some comment. It is stated that"... three months is the maximum time lag for current between the Straits of Florida and the Grand Banks". Considering the paucity of information and the turbulent nature of the flow it is difficult to attach a precise meaning to this statement. It may reasonably be taken to mean that some considerable but unknown fraction of the water particles passing Key West will be found off the Grand Banks in not less than three months. From the records of the Ice Patrol, on the average 87 % of the bergs occur in the period from March to June. If the time from Key West to the Grand Banks is correctly estimated, then anomalies for the months from January through May might be important, with those for March likely to arrive at the height of the iceberg season. All questions of physical significance aside, the correlations for two of these months fail to attain statistical significance at the 0.1 level, one of them is just above the 0.05 level, and the other two are above the 0.001 level. In general, when a number of possible correlations are available it is not very fruitful to fix attention on one impressive one and ignore the others. The "high" correlation for March looks very much like an accident.

An even greater difficulty in accepting the March correlations as meaningful arises from the fact that temperature measurements at Key West are lacking from 1941 through 1945 and the gap has been filled with temperature anomalies measured at Daytona Beach. Since the sample contains only 26 pairs of measures, the interpolated 5 represent 19 % of the data and possible error from this source must be considered. There are 118 months in which temperatures were measured at both places. The standard deviation of the differences between the temperature anomalies at Key West and at Daytona for these 118 pairs is ± 2.21 while the standard deviation of the 21 measures actually available from Key West is ± 2.29. From this it can be inferred that the dispersion of the errors introduced with the Daytona data is probably of about the same size as the dispersion of the Key West measures themselves.

Actually a correlation of 0.65 is not very high. The coefficient of alienation is 0.76 so that the standard error of estimate is reduced only 24 %. For example, if the mean value of the berg counts is used to predict, the standard deviation is ± 360 bergs. If the anomalies having a correlation of 0.65 are used, the standard error of estimate is ± 275 bergs, a reduction of about 24 %, so that the quality of the estimate is still poor.

The uncertainty of the iceberg data has been mentioned. "Surface water temperatures at tide stations, Atlantic Coast, North and South America" says the temperature data "are based upon thermometer readings made in a sample of water drawn by bucket from a foot or two below the surface. The observations are usually made once each weekday at whatever time the observer tends the tide gage..." That the anomalies secured from these data are representative of the fluctuations in the temperature of the entire Florida Current may be doubted so that, even if the whole picture presented by Post were correct, it is questionable whether the numbers he uses do in fact describe the phenomenon about which he wishes to talk. This objection is quite aside from the question of how well the numbers may serve to predict.

From current measurements the time lag for the delayed effect, V, is set at three years and a correlation of − 0.76 is found. If the argument was developed before an inspection of the data, this correlation looks impressive. However, it seems that the time lag of three years inferred from current data is much more doubtful than the estimate of one to three months for the transit from Key West to the Grand Banks. The series of correlations for lags from one to eight years which were presented permit the suspicion that the theory was influenced by the correlations.

We thus arrive at the following conclusions:

1) That the appealingly simple picture of the mechanism connecting iceberg numbers with temperature anomalies at Key West is correct seems quite unlikely.

2) That fluctuation in temperature at Key West is a single cause capable alone of causing the greater part of the fluctuation in iceberg

count is unproven and seems highly unlikely.

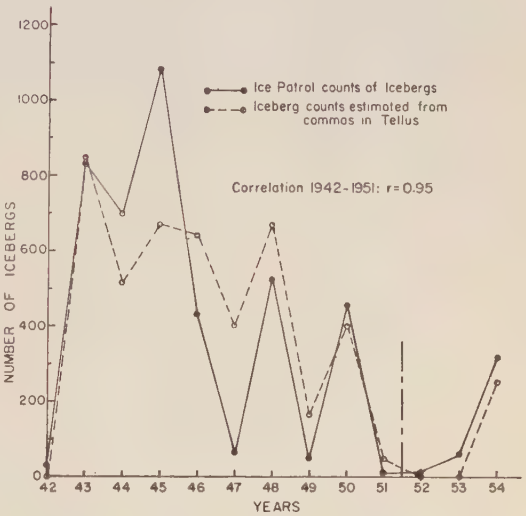
The proposed prediction method fares no better under close scrutiny than the hypothesis. The base of the method is a plot of iceberg counts on a grid of temperature anomalies, one axis being the anomaly of the same year as the iceberg count and the other axis the anomaly of three years before. (Post, 1956, Fig. 6.) There are 23 available iceberg counts. In plotting, the counts have been rounded off as shown in Table II.

Actual International Iceberg Patrol Counts	Actual Counts Accordng To Post's Eigure 6
2	0
2	0
6	0
13	0
22	0
47	0
63	0
30	30
216	225
?	250
430	430
460	475
470	475
475	475
514	500
523	525
575	550
664	670
700	700
840	840
850	850
872	875
1,087	1,075
1,351	1,350

The system on which rounding off is based is not apparent, particularly for the low values. One count too many has been plotted and given a value of 250. The plot made, it is contoured into four zones of approximately equal width, three of which represent a range of 200 bergs while the fourth has a range of 300 bergs. There is no indication of the band to which the dividing numbers 0, 200, 500, and 700 are to be assigned. Inspection shows that the contouring here is extremely unorthodox. Ten out of the 24 counts shown are included in the wrong zones and many of the misplaced counts are not "near misses". This incredible plot seems to represent the entire substance of the prediction method since all subsequent graphs and tables are derived from it. Post's Figure 7 purports to be constructed by reading values from Figure 6 but there are serious disagreements in the values shown in the two figures as was substantiated by the readings made independently by five qualified men.

It is asserted that "a coefficient of +.94 was obtained between computed and observed numbers grouped as in Figure 7 which allows a leeway of ± 100 bergs". The "Iceberg Prediction Table", Post's Table I, says, "For iceberg = 100 correlation coefficient is +94 (see Figure 7)". The process intended is not sufficiently clear to permit a definite check. If one accepts the values shown in Post's Figure 7 and correlates them, the result is 0.83. If the zones as taken in Post's Figure 6 are numbered from 1 to 5 lumping the "No iceberg" region with the 0 to 200 berg zone as 1, the right of assigning borderline counts in the most advantageous way is reserved, and the actual counts are rounded by anything up to ± 100 in such a way as to force the greatest degree of agreement, then the zone number correlation can be forced up to 0.92. Some such purely artificial method which has little relation to the original data seems to have been used.

Verification for the predicting method is claimed for the past three iceberg seasons and *Newsweek Magazine*, July 13, 1953 and *Science Newsletter*, May 1, 1954 and April 30, 1955



are cited, but no criterion by which agreement was judged is offered. As is well known, short runs can be highly misleading. The reader is referred to Figure 1 which shows icebergs as correlated with commas in *Tellus* 8, pages 102—111. The years 1952—1954 were not included in the original analysis and can therefore be considered as "prediction". As can be seen the agreement is much the same as the period 1942—1951. If the comma "prediction" is applied to the iceberg counts for 1937—1941, however, it fails completely, producing an average miss for the five tries of 544 icebergs. Figure 1 should be compared with Post's Figure 7. There is little reason to suppose that Post's predictor will hold up over any long period based as it is on a highly imaginative and unorthodox piece of contouring which fails to "predict" very well even the counts on which it is based.

Conclusion

In conclusion a few cautionary suggestions may be offered on the use of correlations. Correlations may be used freely in the search for possible connections among phenomena and are well defined and meaningful for the samples for which they are calculated. However, inferences from a sample to the population from which it was drawn should be made cautiously with due consideration for the influence of sample size and for the devaluation inherent in selecting one high correlation from a series. Data should never be manipulated to produce a "high" correlation and cause and effect should not be inferred from correlation alone, although correlation may be used as support for other arguments. It would be well if authors would withhold from publication papers whose arguments are constructed to account for a high correlation.

REFERENCES

Cohen, M. R., 1944: *A Preface to Logic.* New York, Henry Holt and Co., pp. 132—134.

Gardiner, M., 1952: *In the Name of Science.* New York, G. P. Putnam's Sons, pp. 173—185, 299—311.

Iselin, Columbus O'D., 1938: "A Promising Theory Concerning the Causes and Results of Long Period Variation in the Strength of the Gulf Stream System." *Trans. Amer. Geophys. Union*, **19**, pp. 234—244.

— 1940: "Preliminary Report on the Long period Variations in the Transport of the Gulf Stream System." *Papers in Physical Oceanography and Meteorology*, **8**, No. 1, pp 1—40.

Poincaré, H., 1905: *Science and Hypothesis.* New York, Dover Publications, Inc., p. 141.

Post, Louis A., 1956: "The Role of the Gulf Stream in the Prediction of Iceberg Distribution in the North Atlantic." *Tellus* **8**, pp. 102—111.

United States Department of Commerce, Coast and Geodetic Survey, "Surface Water Temperatures at Tide Stations, Atlantic Coast, North and South America." Special Publication No. 278, Revised (1951).

Wertheim, G. K., 1953: "Studies of the Electric Potential Between Key West, Florida and Havana, Cuba." *Woods Hole Oceanographic Institution*, Reference No. 53—95 (Unpublished manuscript).

— 1954: "Studies of the Electric Potential Between Key West, Florida and Havana, Cuba, No. II." *Woods Hole Oceanographic Institution*, Reference No. 54—68 (Unpublished manuscript).

Fig. 1: Comparison of iceberg numbers from 1942 to 1954 as counted by the International Ice Patrol and as estimated from commas in *Tellus*. (1942—1951 sample; 1952—1954 "prediction".)

NOTES

Note on the Self-Sustained Oscillations of a Simple Thermal System.

By PIERRE WELANDER, Institute of Theoretical Physics, University of Stockholm, and International Meteorological Institute in Stockholm

(Manuscript received April 30, 1957)

Some years ago the author carried out some convection experiments in which a U-shaped narrow tube filled with water was heated from below and cooled from above. At one occasion, when the ends of the tube were connected with two open water reservoirs (Fig. 1), it was observed that the free surfaces of the reservoirs did raise and lower in a periodic manner, indicating an oscillatory flow through the tube. Repeating the experiment under varying conditions, it was found that the period of this oscillation did change strongly with the dimensions of the tube and the area of the free surfaces. It was also noted that the oscillation did occur only when the area of the free surface surpassed a certain critical value, which depended on the intensity of the heating and cooling. Only by changing the cross-section area of the tube the period of the oscillation could be made to vary from about 20 second to more than 4 hours. In all cases the period was much larger than the corresponding period for free gravity oscillations. Record of a typical oscillation is seen in Fig. 2.

This phenomenon, which does not seem to have described earlier, was recognized as a self-sustained, non-linear oscillation of the kind described in text-books on non-linear mechanics (MINORSKY, 1947). A qualitative explanation of the oscillations is easily given. Suppose that the water is initially at rest and that then, due to the action of some outer disturbance, a small amount of water is displaced, say, from the left to the right shank of the tube. The left shank then gets a small excess of colder water, the right shank a small excess of warmer water, and a buoyancy force appears which drives the flow from the left

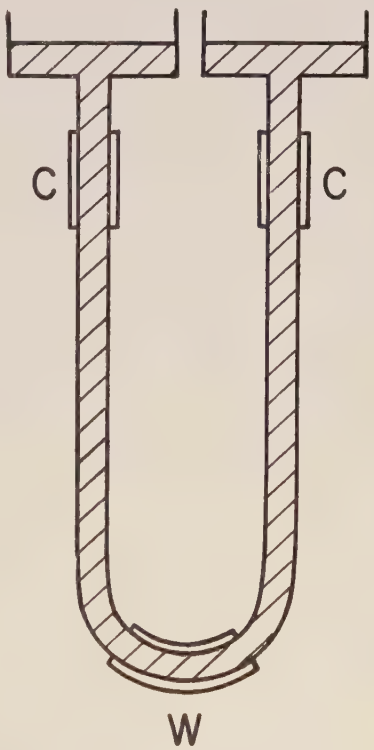


Fig. 1. Scetch of the oscillating system.

to the right shank. As time goes on a difference in level between the reservoirs is built up and slows down the flow. Finally, the buoyancy force and the pressure gradient due to this difference in level balance, and the flow stops. This, however, is not an equilibrium state. When the motion stops, the heating and cooling tend to create a symmetric temperature distribution, thereby reducing the buoyance force, and the flow starts in the opposite direction. On the other hand, when the difference in level is zero the exists still a buoyancy force which sustains the flow, and the motion proceeds periodically.

It can be verified theoretically that no oscillations could occur unless the area of the free surfaces of the reservoirs is sufficiently large. Suppose that due to the initial disturbance a volume dV of water has been displaced from the left to the right shank. If the maximum density difference created by the heating and cooling is $\Delta\varrho$, the difference in pressure between the two shanks due to this displacement cannot exceed $\frac{2g\Delta\varrho dv}{a}$, where a is the cross-section area of the tube and g is the acceleration of gravity. On the other hand, the volume transport creates a difference in level between the reservoirs $\frac{2dv}{A}$, where A is the area of the free surface in one of the reservoirs, and this gives an opposing pressure of magnitude $\frac{2g\varrho dv}{A}$. If oscillations should occur, it is thus necessary that $\frac{a}{A} < \frac{\Delta\varrho}{\varrho}$, or $\frac{a}{A} < 2.\ 10^{-4}\,\Delta\vartheta$, if one introduces the coefficient of thermal expansion for water (at 20°C). If, as example, $\Delta\vartheta = 50°$ one must have $A > 100$ a.

Theoretical expressions may also be derived

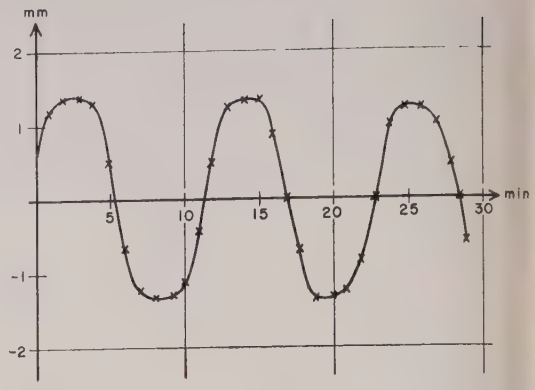


Fig. 2. Typical record of an oscillation. $a = 0.48$ cm² $A = 124$ cm², $\Delta\vartheta \approx 30°$ C.

for the period T of fully developed oscillations. For instance, in the case of large reservoirs and laminar flow through the tube one finds $T = k\,\frac{\nu l A}{g a^4}$ where ν is the kinematic viscosity of water, l is the length of the tube, a and A are the areas defined earlier, and k is a pure number of the order of 1. The observed values of T were found to agree reasonably with the above theoretical result. Some of the theoretical discussions may be presented in a later contribution.

The described experiment may not have any direct application to natural systems since it is an artificially constructed laboratory experiment. Nevertheless, the general information we might collect from this and similar experiments on self-sustained thermal oscillations may provide a base for better understanding of some of the important oscillatory phenomena in the thermal cycles of the earths fluid envelopes, about which we up to now know very little.

REFERENCES

Minorsky, N., 1947: Non Linear Mechanics, Ann Arbor.

Letter to the Editor

The Effect of Fuel Combustion on the Amount of Carbon Dioxide in the Atmosphere

Dear Sir,

Although this matter has been carefully examined in the three valuable papers by H. Craig, R. Revelle —H. Suess, and J. Arnold—E. Anderson, on the exchanges of CO_2 between atmosphere and oceans, in the February issue of *Tellus*',[9] pages 1 to 32, there is one point on which I may be able to throw some further light.

This point is the large discrepancy between recently measured depressions of the radio-carbon activity, (known as the "Suess effect"), and the apparent increase in atmospheric CO_2 as given by measurements of this quantity between 1870 and 1935, (Callendar 1940). Both Revelle and Suess, p. 25, and Arnold and Anderson, p. 31, consider that no increase as large as the 10 % indicated by the latter measurements could actually have occured in such a period, in view of the size of the active CO_2 exchange reservoirs shown by their measurements. However, they suggest other factors which might conceivably account for the discrepancy.

Two suggestions follow which may, in part at least, account for the latter, without having to invoke an unacceptably large error in the atmospheric CO_2 observations made in the 1930^s, or an highly improbable rise in sea temperature.

(1) The possibility that the part of the active CO_2 exchange reservoir which constitutes the land biosphere and humus, may be considerably smaller or slower in action than supposed, is mentioned by Arnold and Anderson. p 31. Two considerations support this:

(a). For active exchange—A large fraction of this carbon is in the heart wood of trees, or below the surface of bogs, etc., and thus inaccessable during short periods.

(b). For Pco_2 control—the considerable amount of CO_2 released by clearance of forests, cultivation and drainage of humic soils, etc., (here estimated at about 10^{17} grams this century), should be enough to balance a rise of several percent in the active part of this reservoir. If so the latter would be ineffective in regard to Pco_2, although active in exchange.

(2) A Meteorological cause is suggested for some of the relatively high CO_2 values observed in the 1930^s: The year 1934, when Prof. Buch took most of his Persamo series, was one of record mildness in N. Europe, the period May to October being as much as 3° C above former averages in the N. Baltic region. This unusual warmth should cause rapid bacterial oxidation of organic residues in the great "muskeg" areas to the east, and also release CO_2 from shallow water bodies in the region leading to high CO_2 in slow moving air currents. (Buch's average CO_2 value for days when continental air reached Petsamo in May to October 1934 was as high as 341 p.p.m.).

An oceanic cause may also be considered, for Brown (1953) has shown that the average surface temperature of the N.E. Atlantic rose by nearly 1° C between the $1920:^s$ and $1930:^s$. This influx of warmer water might lead to some high CO_2 values near the surface, by disturbing the pressure equilibrium of CO_2 saturated cold water bodies, or even by itself having above normal CO_2 tension for these latitudes. If so it could account for the relatively high average value of 319 p.p.m. reported by Prof. Buch on both his crossings of the N. Atlantic in the summers of 1932 and 1935. It seems probable that a combination of these unusual events could give temporary and local averages up to 5 % above the general mean for the zone.

Coming to the present time. The valuable and extensive new series of CO_2 observations reported in Tellus by Fonselius and Koroleff (1955—1957), show certain features which may be related to the meteorological-biological factors mentioned in (2) above. For example; the high CO_2 average of 349 p.p.m. found at the Finnish stations in the warm period July to September 1955, as compared with only 322 for the same months in 1956, which were relatively cool, wet and windy over a

great area of N.W. Europe It seems probable that the meteorological conditions in 1956 would be unfavourable to decay processes, and also to the accumulation of excess CO_2 in surface air currents.

The overall average computed from the new Scandinavian CO_2 measurements for 1955 and 1956 (to october), comes to 326 p.p.m., and is approximately 36 p.p.m. above the best 19th century averages. This is particularly interesting because the amount of CO_2 released by fossil fuel combustion since the turn of the century, (2.8 $10.^{17}$ gms.), is sufficient to cause an increase of 35 p.p.m. in the whole atmosphere. However, the significance of this remarkable coincidence requires much data for adequate discussion, and this cannot be attempted here.

Finally. In view of the probable climate effects of increased atmospheric CO_2, it is interesting, and possibly significant, to note that both the Suess effect, and the temperature rise appears to be considerably less in the southern than in the northern hemisphere. If the former is confirmed, the air exchange between hemispheres will require careful reassessment.

Horsham, April 15, 1957.

Yours faithfully

G. S. Callendar
44 Parsonage Road, Horsham, Sussex, England

REFERENCES

Arnold, J. R., Anderson, E. C., 1957: The Distribution of Carbon-14 in Nature. *Tellus* **9**, p. 28.

Brown, P. R., 1953: Climate fluctuations in the Greenland & Norweg'an seas, 2. *J. R. Met. Soc.* **79**, p. 272.

Callendar, G. S., 1940: Variations of CO_2 in different air currents. *J. R. Met. Soc.* **66**, p. 395².

Craig, H., 1957: The Natural Distribution of Radiocarbon and the Exchange Time of Carbon Dioxide Between Atmosphere and Sea. *Tellus* **9**, p. 1.

Revelle, R., Suess, H. E., 1957: Carbon Dioxide Exchange Between Atmosphere and Ocean, and the Question of an Increase of Atmospheric CO_2 during the Past Decades, *Tellus* **9**, p. 18.

Fonselius, Koroleff: CO_2-values in Scandinavia *Tellus* 7, 8, & **9**.

Current Data on the Chemical Composition of Air and Precipitation X

(For further information see Egnér, H., Eriksson, E., *Tellus* **7**, pp. 134—139, **8**, p. 285 and 517)

Code	mm	mg/m²								pH	$HCO_3 - \frac{\mu val}{l}$	$\varkappa \cdot 10^6 \frac{}{\Omega \cdot cm}$	μg/m³ (= kg/km³)						
		S	Cl	NO_3–N	NH_3–N	Na	K	Mg	Ca				S	Cl	NH_3–N	Na	K	Mg	Ca
	Precipitation February 1957 (D 702)												Air February 1957 (L 702)						
Ri	2	0	32	0	1	18	2	4	8	6.5	*	80	*	*	*	*	*	*	*
Ki	38	10	15	1	0	9	2	3	9	5.9	4	6	*	*	*	*	*	*	*
Ar	16	16	8	2	2	16	3	3	18	6.3	18	13	*	*	*	*	*	*	*
Öj	32	22	9	5	4	8	5	5	28	6.0	7	12	5.5	0.0	2.0	2.9	1.9	1.3	6.3
Rö	0.3	*	3	*	*	1	*	*	*	*	*	*	21.	2.3	2.0	1.0	0.4	1.8	5.0
Of	13	19	5	5	5	5	4	4	9	4.8	0	22	7.5	1.5	1.5	2.6	0.6	1.0	5.4
Br	4	4	7	1	0	6	1	3	3	5.5	*	28	*	*	*	*	*	*	*
ÅF	4	12	7	0	1	19	5	3	10	6.8	*	59	12.	2.4	1.0	0.9	0.5	0.3	2.1
ÅH	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sv	19	14	4	1	0	2	3	4	11	5.4	0	11	—	—	—	—	—	—	—
Rä	16	6	22	5	2	8	7	3	134	7.2	234	57	5.6	4.5	3.4	1.9	0.9	2.7	30.
Åm	28	18	11	6	4	16	4	4	16	5.8	0	11	6.9	0.0	1.5	0.6	0.7	0.6	1.4
Sa	49	46	15	13	23	22	14	6	54	6.4	29	18	13.	0.4	1.7	0.5	1.0	1.1	3.3
Ul	55	37	17	14	10	9	4	4	15	4.6	0	18	17.	0.5	1.2	0.5	0.7	0.5	1.3
Er	58	43	12	11	8	10	5	5	14	4.5	0	21	14.	1.1	1.8	0.8	2.0	2.4	2.1
St	50	46	17	17	7	11	12	7	16	5.6	0	15	18.	0.7	2.0	0.6	0.7	0.7	2.0
Fo	36	24	9	9	3	10	3	6	20	4.7	0	18	5.7	3.7	1.6	1.5	1.0	1.2	3.3
Kv1	59	361	27	16	42	24	21	19	307	4.1	0	84	23.	1.6	3.9	0.6	0.6	2.1	3.1
Kv7	64	63	20	14	13	18	8	7	25	4.4	0	28	7.0	2.2	1.9	0.6	0.5	0.5	3.3
VK	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
La	46	52	95	16	12	44	9	9	32	4.6	0	43	1.1	22.	2.8	3.4	1.8	0.7	9.8
Bo	57	74	312	17	10	193	13	41	28	4.5	0	44	1.7	7.0	1.5	3.0	0.9	1.1	6.3
Vi	56	93	952	24	19	561	30	73	35	4.3	0	94	2.2	1.7	0.2	0.2	0.4	0.0	1.4
Fa	45	51	23	12	10	14	6	7	22	4.8	0	16	0.3	1.4	1.6	0.6	0.5	0.0	2.0
Fl	61	47	59	19	10	42	8	9	37	4.9	0	19	13.	4.7	1.7	1.0	1.1	0.1	3.2
Am	73	66	69	25	17	53	14	14	38	4.8	0	20	—	—	—	—	—	—	—
Si	83	80	89	23	21	61	7	11	28	4.6	0	23	—	—	—	—	—	—	—
Pl	68	104	148	33	56	83	10	21	36	4.4	0	39	15.	4.4	4.6	1.8	1.1	0.7	5.0
Sö	51	53	49	24	10	36	8	9	36	4.5	0	27	—	—	—	—	—	—	—
Sm	33	39	17	14	11	11	4	4	18	4.5	0	26	5.3	2.8	1.6	0.7	0.4	0.0	2.0
Sy	46	64	82	14	20	70	10	8	32	4.9	0	26	4.1	14.	3.0	0.8	0.5	1.1	4.4
BH	48	61	47	20	18	31	6	12	23	4.5	0	26	11.	8.6	3.2	1.3	1.1	2.8	10.
Sk	66	36	123	32	45	64	9	11	34	4.5	0	36	19.	0.0	4.3	1.3	0.9	0.6	3.3
Al	42	80	80	18	23	38	14	10	32	4.4	0	39	10.	2.7	3.1	1.1	0.6	1.1	4.2
Hi	16	25	22	7	6	23	3	4	21	5.6	10	28	—	—	—	—	—	—	—
Ta	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
An	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Yt	17	20	381	1	4	210	12	30	14	6.0	14	96	—	—	—	—	—	—	—
Gj	72	24	584	2	8	385	19	44	20	5.5	0	42	—	—	—	—	—	—	—
Fn	42	21	138	3	9	81	5	10	7	5.8	6	18	—	—	—	—	—	—	—
Få	41	20	40	1	3	29	5	6	6	5.6	0	8	—	—	—	—	—	—	—
Vå	7	17	9	0	1	18	4	3	9	7.2	188	48	4.5	5.9	7.0	0.7	0.9	0.5	5.8
Tr	33	21	7	7	7	4	4	3	6	5.2	0	10	—	—	—	—	—	—	—
Ke	26	0	8	7	6	7	4	4	9	5.1	0	12	—	—	—	—	—	—	—
Sd	81	69	327	12	30	214	12	30	22	5.1	0	28	—	—	—	—	—	—	—
Da	50	22	140	5	6	98	7	10	17	5.6	4	17	—	—	—	—	—	—	—
Ås	54	40	28	14	9	25	6	6	17	4.7	0	19	3.9	0.0	0.9	0.3	0.5	0.0	0.9
Li	86	158	1 626	24	17	1024	57	136	55	4.5	0	99	4.9	23.	1.2	15.	1.3	2.0	2.7
Jy	31	14	26	*	*	14	15	2	22	5.5	0	16	5.5	8.5	5.8	0.5	0	0	8.5
Ka	4	11	7	3	4	6	0	2	25	6.3	0	51	0	30.	6.1	4.0	0	0	19.
Ku	15	15	18	*	*	7	7	1	16	6.6	10	16	0	86.	6.8	0.5	0	0	92.
Pu	39	20	18	6	3	3	0	2	16	5.3	0	9	4.6	16.	4.6	0.4	0	0	0.
So	12	9	8	2	1	8	0	1	7	6.3	0	12	0	0	5.2	0.7	0	0	0.
Tv	37	32	45	10	7	23	3	4	21	6.3	0	18	—	—	—	—	—	—	—

Code	mm	mg/m²								pH	$HCO_3 - \frac{\mu val}{l}$	$\frac{\varkappa \cdot 10^6}{\Omega \cdot cm}$	$\mu g/m^3$ (= kg/km³)						
		S	Cl	NO_3–N	NH_3–N	Na	K	Mg	Ca				S	Cl	NH_3–N	Na	K	Mg	Ca
	Precipitation February 1957 (D 702)												Air February 1957 (L 702)						
Öd	38	56	121	15	20	70	10	13	25	4.7	0	36	11.	21.	2.8	10.	0.7	1.5	2.0
As	75	87	115	21	36	83	10	14	31	4.6	0	28	11.	0.0	2.7	1.3	1.5	0.7	4.6
Ty	34	70	61	18	40	98	7	11	32	6.7	67	38	3.0	5.4	5.1	1.2	0.4	0.7	2.8
Ab	73	54	481	15	1	366	32	46	80	5.8	11	54	4.8	6.9	2.3	7.4	1.9	2.8	11.
Ed	53	54	284	10	4	181	9	26	43	5.0	0	37	13.	6.2	2.6	1.8	0.7	1.3	12.
Le	51	80	201	16	40	56	11	17	49	4.4	0	49	52.	7.5	6.2	1.9	1.5	4.4	15.
Ro	72	98	147	12	16	71	15	15	50	4.4	0	36	31.	16.	4.6	3.7	2.1	1.5	19.
NA	273	226	221	20	70	914	55	131	572	6.4	77	40	17.	4.6	7.2	2.9	1.7	4.8	9.8
Au	97	116	69	3	26	20	7	54	73	5.2	0	16	13.	4.4	5.7	3.1	2.0	7.7	1.7
Ba	128	109	69	19	14	21	13	5	40	5.1	0	10	7.5	4.5	4.0	3.0	1.8	0.4	2.0
Fe	94	60	51	2	—	16	7	16	38	5.4	0	8	1.1	2.8	0	0.1	0.2	0.3	1.5
Sc	62	128	185	20	12	102	13	31	56	4.4	0	48	8.4	20.	8.4	12.	5.0	1.4	4.1
BV	28	68	73	20	27	18	8	23	119	4.7	0	62	18.	10.	5.6	2.6	2.0	1.4	2.1
Ho	48	47	0	0	4	3	10	4	34	5.2	0	11	10.	7.9	6.4	3.7	2.5	1.1	1.1
Bn	70	107	88	22	26	22	11	35	209	5.7	11	34	26.	0.	12.	6.5	4.4	0.6	1.9
BL	165	223	0	23	94	103	10	46	157	6.2	8	16	13.0	11.6	6.1	2.5	2.7	0.7	2.4
U	125	204	45	31	10	97	19	14	88	4.4	0	31	44.8	16.3	4.1	2.5	2.0	1.5	1.4
SA	85	104	31	11	30	80	10	9	36	4.6	0	23	15.5	4.2	4.0	1.2	0.5	0.3	9.8
B	229	190	98	29	26	96	18	13	48	4.7	0	15	18.9	12.0	3.9	7.6	2.3	4.0	4.1
D	107	83	123	17	8	65	8	8	38	4.9	0	14	8.9	2.8	1.9	1.0	0.5	0.5	0.9
DB	55	102	91	17	31	50	7	6	20	4.3	0	39	25.6	4.7	6.2	2.2	1.1	1.0	2.0
	Precipitation March 1957 (D 703)												Air March 1957 (L 703)						
Ri	30	24	53	3	3	34	4	7	14	4.7	0	20	*	*	*	*	*	*	*
Ki	12	6	6	0	0	2	1	3	19	5.3	0	9	0.0	3.4	0.8	3.5	1.1	0.4	3.9
Ar	17	14	2	3	6	3	2	2	12	5.4	0	12	5.4	0.0	1.8	3.7	1.2	0.5	3.7
Öj	26	31	18	5	3	7	7	4	43	5.3	1	17	4.1	2.7	1.9	1.0	0.9	1.7	2.9
Rö	28	29	7	3	5	8	3	3	12	4.5	0	17	*	*	*	*	*	*	*
Of	25	19	0	3	4	3	5	2	8	4.6	0	13	5.5	2.5	2.3	1.3	0.5	0.1	1.2
Br	8	8	2	2	3	8	2	2	12	4.8	0	24	*	*	*	*	*	*	*
ÅF	28	25	18	3	4	17	5	4	13	5.1	0	13	5.4	0.0	1.1	0.6	1.6	0.1	1.7
ÅH	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sv	33	23	5	2	1	4	4	3	17	4.8	0	12	—	—	—	—	—	—	—
Rä	13	36	7	5	2	4	3	2	121	7.1	260	62	6.3	3.9	4.0	0.7	0.7	0.4	7.8
Åm	23	17	12	6	5	11	5	3	25	5.7	6	12	4.2	0.0	2.4	1.2	4.2	0.3	1.7
Sa	14	21	9	5	6	9	4	5	28	5.5	0	25	14.	0.0	2.0	0.7	0.5	0.7	4.9
Ul	15	23	10	7	11	4	3	4	20	4.6	0	32	14.	1.1	2.4	0.5	0.4	0.9	1.4
Er	11	17	9	3	3	10	3	3	14	4.5	0	30	8.8	0.0	1.9	0.5	1.0	1.3	2.5
St	16	26	11	7	7	9	8	5	39	5.9	15	26	11.	3.2	2.6	0.5	0.6	0.4	1.5
Fo	25	26	7	7	6	6	4	5	29	4.8	0	19	6.1	3.2	1.7	1.0	0.7	0.9	3.0
Kv1	28	190	28	9	25	21	15	10	258	5.3	0	84	28.	0.0	3.7	0.6	0.5	3.5	2.5
Kv7	20	41	6	6	11	8	4	4	37	4.2	0	45	11.	1.5	2.8	0.7	0.6	1.1	12.
VK	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
La	20	45	28	13	13	15	8	5	32	4.3	0	45	7.8	3.5	3.5	1.2	0.7	1.4	2.6
Bo	66	60	65	17	18	48	9	16	37	4.6	0	21	6.7	3.5	1.8	1.8	0.4	1.0	3.5
Vi	34	86	387	18	15	280	17	41	29	4.1	0	91	12.	16.	1.6	11.	0.7	3.4	2.6
Fa	36	31	14	9	8	8	5	4	21	4.6	0	19	2.5	0.0	1.4	0.8	0.2	0.9	2.5
Fl	42	36	37	9	7	23	5	6	23	4.4	0	21	9.7	2.7	1.6	0.5	0.4	0.2	2.7
Am	42	49	52	11	10	49	8	9	46	5.3	0	19	—	—	—	—	—	—	—
Si	49	65	59	14	17	44	6	9	38	4.6	0	25	—	—	—	—	—	—	—
Pl	39	94	128	19	42	78	10	13	38	4.0	0	45	13.	1.8	4.0	0.8	0.4	0.4	1.8
Sö	28	45	26	13	14	13	6	6	44	4.6	0	29	—	—	—	—	—	—	—
Sm	12	36	7	6	9	5	2	2	20	4.2	0	41	6.7	2.5	1.9	1.3	1.3	0.8	2.3
Sy	32	32	26	9	9	12	3	4	76	6.2	39	21	7.8	12.	4.1	0.9	0.7	0.7	10.
BH	44	48	30	11	14	18	3	7	26	4.3	0	26	0.0	4.5	2.7	0.6	0.3	3.5	17.
Sk	36	58	65	15	20	40	5	8	40	4.4	0	39	14.	9.6	5.2	6.5	1.1	4.2	3.8

Code	mm	mg/m²								pH	$HCO_3^- - \frac{\mu val}{l}$	$\frac{\varkappa \cdot 10^6}{\Omega \cdot cm}$	$\mu g/m^3$ (= kg/km³)						
		S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca				S	Cl	NH_3-N	Na	K	Mg	Ca
	Precipitation March 1957 (D 703)												Air March 1957 (L 703)						
Al	28	61	44	12	15	22	7	9	55	4.4	0	39	13.	2.9	3.4	0.9	0.8	1.0	2.5
Hi	39	50	52	16	24	34	7	7	54	4.6	0	33	—	—	—	—	—	—	—
Ta	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
An	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Yt	33	34	54	6	14	32	4	13	8	4.6	0	23	—	—	—	—	—	—	—
Gj	61	47	230	7	17	128	6	18	29	5.0	0	24	—	—	—	—	—	—	—
Fn	38	32	11	9	19	9	3	3	32	5.6	3	13	—	—	—	—	—	—	—
Få	36	58	40	45	28	27	11	11	23	4.7	0	29	—	—	—	—	—	—	—
Vå	10	17	8	3	5	36	4	3	17	6.6	79	34	13.	0.0	7.9	0.5	0.6	0.2	11.
Tr	21	16	7	6	7	4	2	2	28	5.3	0	12	—	—	—	—	—	—	—
Ke	13	18	4	6	9	4	3	3	16	5.3	4	22	—	—	—	—	—	—	—
Sd	85	96	180	29	48	105	13	17	37	4.3	0	32	—	—	—	—	—	—	—
Da	51	51	26	16	18	14	8	5	71	5.5	4	17	—	—	—	—	—	—	—
Ås	30	29	18	8	8	14	5	4	24	4.7	0	17	7.8	2.9	3.0	1.1	0.5	0.8	0.8
Li	39	92	432	24	18	282	18	40	41	4.2	0	80	10.	16.	2.0	1.2	1.5	2.1	8.8
Jy	2	0	0	0	0	2	2	2	21	6.9	341	44	4.6	0	0.3	2.0	0.4	0.8	2.0
Ka	7	0	0	0	0	0	3	1	57	6.4	28	18	0	22	2.4	3.0	2.7	0.8	9.4
Ku	6	4	0	1	1	*	6	1	*	6.0	0	12	*	*	1.8	1.0	2.1	1.7	4.7
Pu	10	8	2	12	10	3	5	1	4	5.5	0	12	6.2	360	1.3	0.2	0.7	0.3	1.0
So	2	0	0	2	2	2	2	2	2	6.3	149	62	5.4	65	1.8	1.6	0.6	0.6	0.6
Tv	9	2	3	6	7	0	0	4	0	4.4	*	50	0	66	0	6.8	1.9	1.3	12.6
Öd	21	38	60	11	16	43	8	8	30	4.6	0	42	8.5	3.6	4.6	1.9	1.1	1.1	1.7
As	50	78	124	21	34	81	16	12	40	4.3	0	41	12.	2.1	4.3	8.6	0.9	0.8	4.3
Ty	37	91	100	19	41	122	10	10	65	6.0	40	46	11.	5.2	6.5	3.4	0.9	0.8	4.5
Ab	59	71	258	17	5	166	19	27	68	5.0	0	35	11.	9.5	5.1	3.3	4.2	18.	24.
Ed	40	38	60	7	10	51	12	12	43	5.9	5	23	18.	6.4	2.7	0.8	1.1	0.9	8.1
Le	36	101	212	20	44	58	13	25	58	4.0	0	87	21.	18.	8.2	2.5	1.9	3.4	3.3
Ro	26	75	75	14	17	33	16	10	68	4.2	0	63	30.	14.	7.5	4.3	3.1	3.5	16.
NA	133	115	446	18	89	308	31	44	238	6.3	61	34	0.0	9.9	9.7	5.2	1.7	2.0	3.9
Au	40	47	44	19	27	4	9	21	95	5.8	16	25	10.0	3.7	6.3	1.3	1.7	8.5	2.2
Ba	53	20	0	11	23	5	4	2	41	5.4	0	9	3.9	3.1	3.4	0.1	0.4	1.1	2.0
Fe	54	43	5	18	23	4	3	4	33	5.4	0	9	4.9	5.3	3.8	3.9	3.6	3.2	5.2
Sc	41	52	183	21	17	84	15	15	122	4.7	0	55	11.3	33.4	9.3	16.2	4.3	3.0	9.4
BV	85	93	99	32	47	14	15	42	216	4.9	0	34	26.4	6.2	5.6	1.0	1.0	0.6	2.5
Ho	46	29	46	15	21	0	6	6	71	6.2	20	14	8.3	10.2	—	0.9	0.7	0.8	2.5
Bn	76	73	60	50	39	14	17	12	262	5.5	0	32	10.4	17.5	8.7	1.2	1.7	3.1	3.8
He	32	20	63	30	26	27	51	5	73	6.4	22	29	—	—	—	—	—	—	—
Lz	107	198	135	0	35	37	32	36	138	5.4	0	21	14.4	9.0	7.1	6.4	6.8	*	48.5
BL	30	53	24	10	43	8	13	18	146	5.0	0	61	9.4	1.9	9.1	0.9	0.9	0.9	2.4
U	42	104	26	19	50	29	9	9	90	5.0	0	39	36.1	4.4	5.4	0.9	0.9	0.7	0.9
SA	37	93	32	14	36	33	7	6	54	4.3	0	40	12.4	1.1	8.1	0.9	0.8	0.5	0.9
B	126	123	42	40	63	31	18	9	52	4.7	0	18	12.8	4.7	6.1	1.4	2.4	0.8	1.7
D	22	43	33	12	11	25	7	5	50	5.5	0	31	9.5	7.	3.8	0.5	0.	0.5	0.8
DB	44	152	132	25	75	75	35	10	78	5.0	0	49	34.4	5.2	7.	1.3	0.	1.2	1.7
	Precipitation April 1957 (D 704)												Air April 1957 (L 704)						
Ri	42	35	18	5	6	20	3	7	12	4.7	0	23	*	*	*	*	*	*	*
Ki	7	4	1	0	0	4	2	1	5	5.5	*	19	2.1	6.8	1.0	6.6	0.6	0.8	3.3
Ar	3	9	3	0	2	3	2	2	17	6.1	*	37	3.9	0.2	1.2	1.3	0.4	0.3	8.4
Öj	15	18	5	1	2	3	4	3	33	6.3	30	19	6.7	6.2	2.9	1.1	1.0	0.5	4.9
Rö	6	7	3	1	0	5	2	2	15	5.1	0	30	*	*	*	*	*	*	*
Of	9	5	4	1	2	4	3	3	22	5.8	17	24	0.0	6.5	2.3	0.5	0.3	0.3	4.8
Br	3	1	6	1	1	10	3	2	14	6.9	*	48	1.9	3.6	0.9	0.6	0.6	0.1	3.0
ÅF	34	0	12	1	0	16	4	4	22	6.4	14	9	—	—	—	—	—	—	—
ÅH	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Code	mm	mg/m²								pH	HCO_3 $\frac{\mu val}{l}$	$\varkappa \cdot 10^6$ $\frac{}{\Omega \cdot cm}$	μg/m³ (= kg/km³)						
		S	Cl	NO_3—N	NH_3—N	Na	K	Mg	Ca				S	Cl	NH_3—N	Na	K	Mg	Ca
	Precipitation April 1957 (D 704)												Air April 1957 (L 704)						
Sv	23	9	4	1	0	2	4	2	18	6.2	18	8	—	—	—	—	—	—	—
Rä	21	6	15	3	3	4	4	3	183	7.2	396	57	3.4	4.0	1.9	0.3	0.5	0.3	4.2
Åm	18	16	11	3	4	10	5	4	15	6.3	30	15	5.5	4.7	2.0	0.6	0.6	1.4	8.0
Sa	10	25	8	3	3	9	4	6	27	6.1	17	35	8.7	5.3	2.0	0.6	0.7	1.6	6.2
Ul	11	19	3	3	5	4	3	3	19	5.0	0	32	4.2	2.1	2.0	0.2	0.3	0.0	0.6
Er	9	21	6	2	2	7	4	4	20	5.3	0	33	6.2	0.0	1.6	0.8	1.4	0.2	1.9
St	11	24	8	4	2	9	6	4	32	6.1	40	32	7.6	3.6	2.3	0.3	0.6	0.1	1.1
Fo	38	32	3	5	8	5	5	4	21	5.2	0	14	2.7	0.0	1.4	0.6	0.4	0.9	6.2
Kv1	13	197	30	6	25	11	9	11	301	6.1	18	170	29.	1.4	5.0	0.5	0.9	3.4	3.7
Kv7	13	27	8	4	7	6	4	3	20	5.0	0	35	—	—	—	—	—	—	—
VK	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
La	19	26	11	5	8	6	5	3	20	5.4	0	21	0.0	7.1	2.9	0.9	2.2	0.6	3.0
Bo	26	36	32	7	8	35	5	9	30	6.1	16	24	*	*	*	*	*	*	*
Vi	19	40	200	6	4	124	9	22	15	4.5	0	82	8.9	40.	1.2	3.6	0.5	2.0	1.5
Fa	40	40	8	4	0	10	4	4	30	5.1	0	16	*	*	*	*	*	*	*
Fl	34	36	8	3	9	6	7	3	16	5.5	0	13	5.8	3.4	1.5	0.6	0.7	0.8	1.3
Am	32	34	10	0	0	20	6	6	29	6.7	58	16	—	—	—	—	—	—	—
Si	19	29	29	5	3	22	7	5	24	6.1	6	24	—	—	—	—	—	—	—
Pl	26	55	36	18	20	21	6	6	29	4.5	0	44	5.2	4.1	4.8	0.5	0.3	0.9	1.6
Sö	28	51	11	5	6	12	3	5	63	6.1	17	25	—	—	—	—	—	—	—
Sm	57	59	16	8	22	8	5	4	24	4.8	0	20	5.0	3.9	2.5	0.4	0.6	3.2	10.2
Sy	21	28	10	4	5	7	4	3	37	6.2	33	20	1.7	9.0	2.7	0.7	0.6	2.0	9.9
BH	19	37	17	6	15	8	8	4	22	5.7	50	25	2.7	0.0	2.5	0.2	0.2	0.3	6.0
Sk	19	39	26	8	17	20	5	4	26	5.1	0	34	15.	6.3	4.1	0.9	0.6	0.9	7.0
Al	14	45	25	7	11	15	9	6	43	4.8	0	49	6.4	0.8	2.9	0.4	0.3	0.4	3.7
Hi	10	16	8	2	7	7	5	2	14	5.9	50	24	—	—	—	—	—	—	—
Ta	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
An	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Yt	37	35	187	4	10	120	5	17	13	5.8	15	29	—	—	—	—	—	—	—
Gj	57	50	188	4	13	135	8	18	20	5.7	6	22	—	—	—	—	—	—	—
Fn	15	18	4	4	8	8	2	3	12	6.1	28	18	—	—	—	—	—	—	—
Få	45	41	10	12	20	56	18	6	20	5.8	8	17	—	—	—	—	—	—	—
Vå	3	9	10	1	1	14	2	2	9	7.0	*	60	4.2	0.0	6.8	0.4	1.0	3.1	1.7
Tr	8	11	3	2	3	3	5	2	10	6.3	29	20	—	—	—	—	—	—	—
Ke	10	12	4	2	5	5	3	3	12	6.2	2	20	—	—	—	—	—	—	—
Sd	87	117	227	20	31	139	14	24	31	4.7	0	34	—	—	—	—	—	—	—
Da	14	16	2	3	3	2	2	2	39	6.8	95	24	—	—	—	—	—	—	—
Ås	26	42	13	9	15	18	5	4	21	5.6	0	31	8.8	0.0	2.4	0.6	0.8	0.4	0.3
Li	17	37	161	12	9	89	7	16	19	4.1	0	86	7.2	30.	1.7	18.	1.0	1.4	4.2
Jy	8	14	6	2	2	7	2	2	42	6.9	95	40	0	8.2	1.4	0.5	1.0	0.7	6.3
Ka	6	4	48	2	2	4	63	2	21	6.7	77	62	2.0	38.	2.7	1.7	1.7	0	17.
Ku	12	11	6	2	3	4	1	3	25	6.4	0	22	0	43.	1.7	0.6	1.3	0	46.
Pu	42	21	11	5	6	6	3	3	27	6.1	20	9	1.1	8.9	1.3	1.2	0.6	0	8.4
So	18	19	95	1	3	23	45	3	26	6.3	46	39	0	11.	0.8	0.5	1.3	0	11.
Tv	18	19	10	2	2	6	2	3	27	5.9	52	18	0	9.5	0.3	1.7	1.7	1.1	45.
Öd	9	27	31	4	8	27	4	5	25	5.5	14	53	2.7	3.2	4.4	1.3	0.9	0.4	1.9
As	16	45	19	10	20	32	9	7	33	5.0	0	53	—	—	—	—	—	—	—
Ty	12	38	28	7	21	25	4	7	25	6.4	100	53	9.4	6.7	6.3	2.4	1.1	1.2	3.2
Ab	25	50	171	5	0	146	10	20	62	6.7	74	57	18.	2.0	3.5	3.7	1.2	4.2	24.
Ed	24	43	37	6	8	47	16	10	49	6.3	40	37	14.	0.0	2.4	0.9	0.6	1.0	5.7
Le	5	72	111	18	25	46	11	27	106	5.9	—	390	24.	25.	10.	7.3	1.9	5.8	18.
Ro	6	55	88	13	14	56	10	8	63	5.2	0	16	20.	7.1	5.4	2.2	1.1	2.4	7.8
NA	5	50	55	10	20	40	8	10	47	6.2	—	17	25.	35.	10.	18.	3.1	4.7	9.4
Au	37	178	70	49	81	19	19	30	76	5.2	0	56	18.7	8.1	5.6	2.5	1.5	8.0	2.1
Ba	33	74	37	13	34	10	7	22	39	4.7	0	68	17.7	13.1	12.3	6.5	3.1	6.9	4.5
Fe	43	117	75	26	42	17	17	23	65	4.7	0	40	10.1	9.7	4.9	3.6	5.4	3.1	5.6
Sc	12	60	42	5	11	23	4	8	27	4.5	0	35	14.9	17.9	6.3	7.3	3.4	2.9	5.5

Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3	$\frac{\varkappa \cdot 10^6}{\Omega \cdot cm}$	$\mu g/m^3$ (= kg/km³)						
		S	Cl	NO_3—N	NH_3—N	Na	K	Mg	Ca				S	Cl	NH_3—N	Na	K	Mg	Ca
	Precipitation April 1957 (D 704)												Air April 1957 (L 704)						
BV	18	112	44	24	33	18	7	16	72	5.7	0	68	18.4	6.3	4.5	1.3	1.1	2.8	2.0
Ho	51	86	6	10	29	15	10	27	61	5.0	0	27	12.0	3.2	5.4	2.0	1.3	1.8	1.4
Bn	18	*	140	36	35	37	39	49	478	6.8	193	292	32.5	6.2	11.3	2.3	2.9	3.7	5.1
He	25	103	53	15	27	30	20	18	63	5.3	0	50	—	—	—	—	—	—	—
Lz	77	201	108	16	75	92	15	28	92	6.1	30	27	15.6	6.2	6.9	9.5	3.3	*	7.4
BL	24	147	126	0	33	6	21	*	81	*	*	*	19.0	13.5	11.1	1.9	1.1	*	2.2

— No sampling. * Sample discarded. Stations Ri—Li, Öd—NA analyzed at R. Agric. College, Uppsala; stations So—Tv at Institute of Marine Research, Helsinki; stations Au—Bn at Staatl. Landw. Versuchs- u. Forschungsanstalt, Grötzingen, Baden; stations U—DB at Institut d'Hygiène et d'Épidémiologie, Bruxelles.

G. Brodin, Uppsala — F. Koroleff, Helsinki — H Riehm, E. Quellmalz, Grötzingen, Baden — I. Bouquiaux, A. Mertens, Bruxelles

Coordinates of New Sampling Stations for Air Precipitation and CO_2 from 1 Nov. 1956.

Station	Code	Altitude m	Latitude °N	Longitude °E
Tana	Ta	9	70.4	28.2
Andenes	An	5	69.3	16.1
Ytterøy	Yt	74	63.8	11.2
Giermundnes	Gj	51	62.6	7.2
Fortun	Fn	27	61.5	7.7
Fanaråken	Få	2026	61.5	7.9
Trysil	Tr	362	61.3	12.3
Kise	Ke	128	60.8	10.1
Stend	Sd	—	60.3	5.3
Dalen	Da	77	59.5	8.0

Only Tana and Stend take CO_2 samples. Stend replaces Bergen from 27th of June. Coordinates of the stations Heitersheim He Linz Lz and Bern—Liebefeld BL will be given in the next number of Tellus.

CO_2-values in Scandinavia Febr.—April 1957.

(Cf. Fonselius, Koroleff: Tellus **7**, pp. 258—265)

Station	February					March					April				
	Date	°C	Wind	Weather	CO_2 ml/10 l.	Date	°C	Wind	Weather	CO_2 ml/10 l.	Date	°C	Wind	Weather	CO_2 ml/10 l.
Ab	1	—10.0	S 2	2	3.15						1	0.0	NNW 2	4	3.52
	10	—22.0	calm	0	3.23						10	— 3.0	SSE 1	4	3.19
	20	—12.2	calm	8	3.23	20	−18.1	N 1	0	3.50	20	3.0	SW 7	4	3.14
Öj	2	— 1.6	SW 2	8	3.34	1	− 6.2	SE 2	⚹° 8	3.25	1	7.0	calm	8	3.30
	10	— 7.8	NW 2	7	3.24	10	10.2	calm	0	3.05	10	— 3.1	NW 9	⚹ ▽ 4	3.34
	20	— 4.2	NE 2	8	3.15										

Station	February					March					April				
	Date	°C	Wind	Weather	CO_2 ml/10 l.	Date	°C	Wind	Weather	CO_2 ml/10 l.	Date	°C	Wind	Weather	CO_2 ml/10 l.
Br	1	— 2.0	S 10	⚹ 8	3.24	1	– 7.0	E 1	8	3.25	1	10.0	W 10	1	3.17
	10	— 7.0	E 10	⚹ 8	3.12	10	4.0	W 2	1	3.23	10	— 4.0	N 10	4	3.10
	20	— 4.0	N 1	⚹ 8	3.19	20	– 8.0	SE 15	⚹ 8	3.28					
Ul	1	2.1	SSW 6	10	3.09	1	– 0.9	calm	8	3.37	1	10.5	calm	0	3.20
	10	— 2.1	E 4	10	3.27						10	— 1.5	E 8	8	3.06
	20	— 1.5	calm	4	3.10	20	– 4.8	SE 7	8	3.29					
Fl	1	1.2	SE 3	≡ 10	3.04	1	– 0.4	SW 2	7	3.32					
	10	3.6	W 2	0	2.98	11	4.6	SW 2	5	3.29	10	4.6	SW 2	● 10	3.18
	20	0.0	S 2	6	3.33										
Pl	1	1.0	NE 3	10	3.09	1	2.0	calm	3	3.17	1	10.0	calm	0	3.57
	11	2.0	W 1	9	3.10	11	2.0	SW 2	6	3.32	10	2.0	NW 4	4	3.13
	22	1.0	calm	⚹ 8	3.26	22	6.0	SW 6	0	3.10	22	7.0	calm	10	3.01
Ka	1	— 0.7	S 6	9	3.28	1	– 7.4	SSE 6	0	3.37	1	8.0	SSW 1	0	3.30
	10	—11.8	SSE 1	0	3.64	10	1.0	WSW 4	4	3.46	10	— 3.0	NNE 5	0	3.31
	20	— 3.5	N 6	⚹	3.45	20	–14.0	calm	0	3.21	20	2.8	NE 2	0	3.35
Ri	1	— 4.1	SSE 3	9	3.31	1	– 7 2	SE 2	4	3.26	1	5.6	SE 1	0	3.23
	10	—12.9	W 1	0	3.36	10	– 2.2	SW 2	0	3.34	10	— 1.2	NW 5	4	3.17
	20	— 2.9	N 6	⚹	3.28	20	–10.2	W 1	0	3.34	20	2.0	SE 2	≃	3.17
Lu	1	— 2.3	SSW 3	9	3.16	1	– 6.1	SE 1	0	3.16	1	8.2	WSW 2	4	3.34
	10	—10.5	NW 2	0	3 20	10	– 2.9	WSW 3	0	3.40	10	— 1.4	NNE 3	4	3.28
	20	— 4.0	NNE 7	⚹	3.24	20	–11.2	WNW 2	0	3.20					
Tv	1	1.0	S 8	≡	3.28	1	– 1.0	calm		3.10	1	3.0	SW 2	0	3.16
	10	— 3.5	E 3	4	3.00	10	3.0	SW 3	4	3.28	10	— 1.0	S 1	0	3.33
	20	— 1.0	NW 6	0	3.09	20	– 8.0	SE 1	4	3.31	20	4.0	SW 2	0	3.27
So	1	— 3.4	SSW 4	⚹	3.59	1	– 7.5	S 8	0	3.11	1	— 2.6	SSW 4	0	3.23
	10	— 7.7	NNW 2	4	3.19						10	— 5.8	NNE 3	0	3.13
	20	— 3.8	NE 2	⚹	2.98	20	– 7.0	NW 3	0	3.24	20	2.2	SSW 4	0	3.27
Pu	1	— 5.2	SW 2	0	3.30	1	–21.0	N 1	0	3.20	1	2.9	N 2	0	3.43
	10	—22.0	W 1	9	3.18	10	– 5.2	SSW 4	0	3.31	10	— 0.2	NW 5	4	3.23
	20	— 1.8	E 3	⚹	3.29	20	–20.6	W 2	0	3.30	20	2.9	SE 3		3.21
Bo	1	2.7	N 5	=°10	3.07	1	3.5	N 1	●° 10	3.24	1	4.5	NNE 4	=●°10	3.88
	10	— 6.0	N 4	0	2.97	10	4.7	NNE 1	=●°10	3.47	10	— 0.3	N 4	1	3.34
	20	— 1.8	N 2	=°10	3.15	20	– 5.9	N 5	0	3.13	20	8.2	NNE 4	=° 3	3.17
Vå	1	1.3	NNE 2	1	3.08	1	– 6.4	NNE 5	4	3.11	1	3.4	NNE 5	0	3.80
					4	10	– 0.1	N 2	8	3.42	10	2.4	NNE 2	8	3.80
	20	— 9.0	NNE 2	4	3.2	20	– 7.2	N 24	⚹	3.07	20	8.0	NNE 9	4	3.15
Be	1	6.9	SE 8	8	3.20	1	4.2	SE 8	8	3.34	1	10.1	calm	0	3.51
	10	3.4	NW 4	8	3.33	10	8.2	SE 6	● 8	3.26	10	2.0	WNW 2	⚹ 4	3.20
	20	3.3	SE 4	4	3.28	20	8.7	SSE 6	8	3.20					
Öd	1	6.2	S 3	●° 10	3.23	1	3.1	S 4	2	3.15	1	3.8	calm	≡ 10	3.18
	10	4.3	SW 4	7	3.07	10	4.4	SE 4	0	3.18					
	20	0.4	S 2	10	3.35	20	8.5	SSW 2	● ▽=10	3.11					
As	1	7.2	SW 3	●	3.16	1	3.4	S 3	0	3.19	1	3.8	ESE 2	≡ 8	3.33
	10	0.2	WSW 2		3.06	10	5.9	SSE 4	= 7	3.09	10	6.1	NW 5	8	2.99
	20	0.5	ESE 2	8	3.29	20	9.1	SW 5	● 8	3.49	21	12.1	ESE 4	4	3.16

Symbols used for description of weather conditions, are those, internationally used. The numbers in the weather-column give the cloudiness and the wind force in Beauforts.

S. Fonselius
Intern. Inst. of Meteor., Stockholm

F. Koroleff
Inst. of Marine Research, Helsinki

Esselte, Sthlm 57